M 332 Unit 13

THE OPEN UNIVERSITY
Complex Analysis
Mathematics : A Third Level Course

Unit 13 Properties of Analytic Functions

[redacted]

ASF
62 P

COMPLEX ANALYSIS

P/£ 2-30

THE OPEN UNIVERSITY

Mathematics: A Third Level Course

Complex Analysis   Unit 13

Properties of Analytic Functions

Prepared by the Course Team

The Library
Exeter College
Oxford

WITHDRAWN

The Open University Press

The Open University Press, Walton Hall, Milton Keynes, MK7 6AA.

First published 1975.

Copyright © 1975 The Open University.

All rights reserved. No part of this work may be reproduced in any form, by mimeograph or any other means, without permission in writing from the publishers.

Produced in Great Britain by
Technical Filmsetters Europe Limited, 76 Great Bridgewater Street, Manchester M1 5JY.

ISBN 0 335 05554 0

This text forms part of the correspondence element of an Open University Third Level Course. The complete list of units in the course is given at the end of this text.

For general availability of supporting material referred to in this text, please write to the Director of Marketing, The Open University, P.O. Box 81, Milton Keynes, MK7 6AT.

Further information on Open University courses may be obtained from The Admissions Office, The Open University, P.O. Box 48, Milton Keynes, MK7 6AB.

1.1

CONTENTS

Conventions . . . 4
13.0 **Introduction** . . . 5
13.1 **Local Properties of Analytic Transformations** . . . 8
13.2 **Problems** . . . 15
13.3 **Bilinear Transformations** . . . 22
13.4 **Problems** . . . 37
13.5 **The Point at Infinity** . . . 45
13.6 **Problems** . . . 54
13.7 **Global Mapping Theorems** . . . 58
13.8 **Problems** . . . 67
13.9 **Non-Euclidean Geometry** . . . 70
13.10 **Problems** . . . 78
Notation . . . 81
Index . . . 83

Conventions

Before working through this text make sure you have read *A Guide to the Course: Complex Analysis.*

References to units of other Open University courses in mathematics take the form:

Unit M100 13, Integration II.

The set book for the course M231, Analysis, is M. Spivak, *Calculus*, paperback edition (W. A. Benjamin/Addison-Wesley, 1973). This is referred to as:

Spivak.

Optional Material

This course has been designed so that it is possible to make minor changes to the content in the light of experience. You should therefore consult the supplementary material to discover which sections of this text are not part of the course in the current academic year.

Reminder: This unit is one of the last four units in the course. You are not required to study all of these four units, so you should consult the supplementary material to see what choice is available to you.

13.0 INTRODUCTION

The main emphasis in this course has been on the investigation of complex functions from a purely analytic point of view. Even so, we have been unable to prevent some geometrical ideas from creeping in. The most obvious example is *Unit 2, Continuous Functions*, which is full of pictures and picturable ideas. The notion of a set being connected, and in particular the result (of Problem 1 of Section 2.8) that if f is a function continuous on an open connected set R, and $f(R)$ is open, then $f(R)$ is connected, will serve as instances of what we have in mind. Another example is taken from *Unit 3, Differentiation*. In that unit we began a discussion of a geometrical property of a complex function with a non-zero derivative, which we shall take up again in the present unit. The property is that such a function preserves angles. We also gave some examples in which the images of the net of coordinate lines were used to depict the effects of functions. These various results and ideas should give you an impression of what the present unit is about, because in it we shall pick up these threads again.

In this unit, then, we adopt a more geometrical point of view than we have usually taken in previous units. Accordingly, we shall be thinking of a complex function as a transformation, or mapping. The use of these words is supposed to conjure up in your mind an image something like this. The plane is a large flat sheet of rubbery material. If you pull it about, distorting it in any way you wish, but still finishing up with something flat, then you are making a transformation of the plane. If you draw on the rubber sheet some figures—triangles, circles, open sets, what you will—then after transformation you will see figures which are in all probability grievously distorted versions of their originals. One of our main tasks is to attempt to determine whether there is any property of a figure which survives such a distortion, if the transformation you carry out corresponds to an *analytic* function.

The transformation which corresponds to a complex function f is described in terms of the rubber sheet as follows: take the point z between your finger and thumb and pull it over to the point $f(z)$. (You will need sufficient fingers and thumbs to do this to all points of the plane simultaneously.) This idea is clear enough, surely, to make a strict distinction between a function, and the corresponding transformation, merely pedantic. So you will find that in this unit we use the words function, transformation and mapping more or less interchangeably. You may consequently find adjectives used to qualify words they are not normally associated with: for example, an "analytic transformation" is a transformation corresponding to an analytic function.

The basic concerns of the unit are two kinds of question, and one special kind of transformation. The questions are firstly local questions—what can one say about the behaviour of analytic transformations in a neighbourhood of a point—and secondly global questions—what can one say about the behaviour of analytic transformations in the large? These topics will concern us in Sections 13.1 and 13.7 respectively. The special kind of transformation is the *bilinear transformation*, whose importance in this sort of investigation should become clear to you as you read the unit. We begin the discussion of bilinear transformations in Section 13.3. These topics might seem a trifle humdrum by themselves, but they lead on to a couple of rather more exotic matters. The first is an examination of the notion of a *point at infinity*; and the second a description of a model of a *Non-Euclidean geometry*.

Objectives

After working through this unit you should be able to:

(i) recall and employ the statements of Theorems 1, 2, 3, 4 and 5 and Corollary 2 to Theorem 3;

(ii) prove Theorem 1;

(iii) define the terms *conformal mapping* and *open mapping*, and understand their use;

(iv) either
use Lagrange's Formula to find the Taylor series of the inverse of a function;
or
use Picard's method to find a series which approximates the inverse of a function locally;

(v) define the term *bilinear transformation* and understand its use;

(vi) establish the basic properties of bilinear transformations;

(vii) express a bilinear transformation as a composition of a translation, rotation about the origin, magnification and inversion in the unit circle;

(viii) determine whether $Az\bar{z} + \bar{B}z + B\bar{z} + C = 0$ represents a circle, a straight line or neither;

(ix) determine images of given straight lines and circles under bilinear transformations;

(x) determine images of suitable regions under bilinear transformations;

(xi) construct bilinear transformations by composing appropriate functions so as to achieve a specified transformation;

(xii) define the terms *point at infinity*, *neighbourhood of the point at infinity*, *extended complex plane*, *function defined on* $\hat{\mathbf{C}}$, *continuous extension of a function to* $\hat{\mathbf{C}}$, *singularity of a function at* ∞, *limit of a function near a point* $a \in \hat{\mathbf{C}}$, *continuity of a function at a point* $a \in \hat{\mathbf{C}}$, and understand their use;

(xiii) interpret $f(a) = \infty$, $f(\infty) = a$, $f(\infty) = \infty$;

(xiv) state Theorem 6 and prove it;

(xv) describe the stereographic projection of the complex plane and explain its use in giving a model of the extended complex plane;

(xvi) explain the phrase "a bilinear transformation maps circles to circles";

(xvii) determine the nature of the singularity at ∞ of a function from its Laurent series;

(xviii) define the term *analytic equivalence* and understand its use;

(xix) recall the statements of Theorems 7, 8 and 9, and outline a method of proof of Theorem 9 involving the lemma on page 61;

(xx) construct analytic equivalences (by composition where necessary);

(xxi) outline briefly the historical development of the concept of non-Euclidean geometries;

(xxii) describe the extended complex plane model of Riemann's (non-Euclidean) Geometry, explaining the role of bilinear transformations;

(xxiii) describe the open unit disc model of Lobachevskian geometry, explaining the role of analytic equivalences;

(xxiv) calculate "lengths" of "straight lines" between two "points" in Lobachevskian geometry and Riemann's geometry.

Results Needed from Earlier Units

Unit 2: Open, closed, compact sets.
The Boundedness Theorem (Theorem 12).
Problem 3 of Section 2.2.
Problem 1 of Section 2.8.

Unit 3: Analytic functions preserve angles.
Theorem 7.
Theorem 14.
Theorem 17.

Unit 4: Inside and outside of a contour.

Unit 5: Liouville's Theorem (Theorem 10).
Fundamental Theorem of Algebra (Theorem 11).
$\mathop{\mathrm{Lim}}\limits_{z \to \infty} f(z) = l$.

Unit 6: Corollary 1 to Theorem 15.
l'Hôpital's Rule.

Unit 8: Corollary to Theorem 1.
Theorem 4.
Casorati–Weierstrass Theorem (Theorem 5).
$\mathop{\mathrm{Lim}}\limits_{z \to \alpha} f(z) = \infty$.

Unit 9: Cauchy's Theorem (Theorem 5).
Cauchy's Formula (Theorem 7).
Maximum and Minimum Principles.
Winding number (Section 9.5).
Problem 3 of Section 9.6.

Unit 10: Principle of the Argument.

Unit 11: Uniform convergence of a sequence of complex functions; the Weierstrass M-test.

Television

There is a television programme associated with this unit.

13.1 LOCAL PROPERTIES OF ANALYTIC TRANSFORMATIONS

Suppose that f is a complex function which is analytic on some neighbourhood of a point α, say. Are there any interesting geometrical or topological properties which f must have as a consequence of the assumption that it is analytic?

* * * * * * * *

We have already described one geometrical consequence of analyticity in *Unit 3*, where (on page 111) we stated that

> analytic functions preserve angles (at points at which their derivatives are non-zero).

We shall now examine this property a little more closely than we did in *Unit 3*. To do so we must first define what we mean by the tangent to an arc.

Let γ_1 be a smooth arc passing through α, such that $\alpha = \gamma_1(0)$; then the arc $\tau : t \longrightarrow \alpha + t \cdot \gamma_1'(0), t \in [0, \infty)$, is called the **tangent to** γ_1 **at** α. Note that $\tau(0) = \gamma_1(0)$ and $\tau'(0) = \gamma_1'(0)$. If γ_2 is another smooth arc through α (and $\gamma_2(0) = \alpha$) we call the angle through which the tangent to γ_1 at α must turn in a *counter-clockwise* direction until it coincides with the tangent to γ_2 at α the **angle between** γ_1 **and** γ_2 **at** α. The angle between γ_1 and γ_2 at α is equal to $\operatorname{Arg} \gamma_2'(0) - \operatorname{Arg} \gamma_1'(0)$, up to multiples of 2π, and lies between 0 (inclusive) and 2π. This statement is not quite obvious perhaps, but if you look at the diagrams in Fig. 1, and other similar ones, you should be convinced.

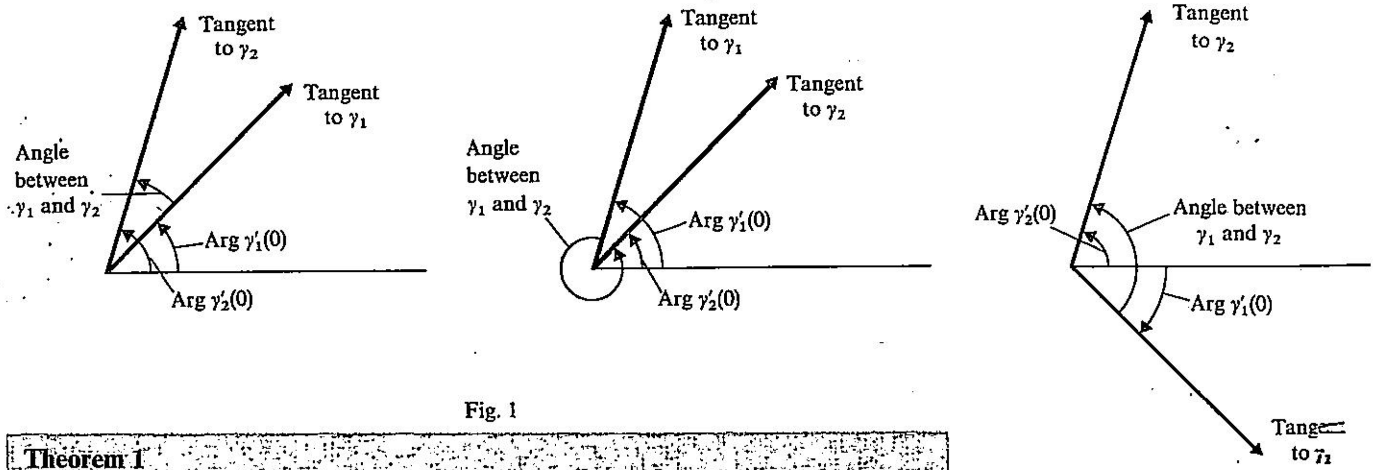

Tangent to γ_2
Tangent to γ_1
Angle between γ_1 and γ_2
Arg $\gamma_1'(0)$
Arg $\gamma_2'(0)$
Tangent to γ_1
Tangent to γ_2
Angle between γ_1 and γ_2
Arg $\gamma_1'(0)$
Arg $\gamma_2'(0)$
Tangent to γ_2
Arg $\gamma_2'(0)$
Angle between γ_1 and γ_2
Arg $\gamma_1'(0)$
Tange
to γ_1

Fig. 1

Theorem 1

If f is analytic on a neighbourhood of α and $f'(\alpha) \neq 0$, then for any pair of smooth arcs γ_1 and γ_2 through α, $f \circ \gamma_1$ and $f \circ \gamma_2$ are smooth arcs through $f(\alpha)$ such that the angle between $f \circ \gamma_1$ and $f \circ \gamma_2$ at $f(\alpha)$ is the same as the angle between γ_1 and γ_2 at α.

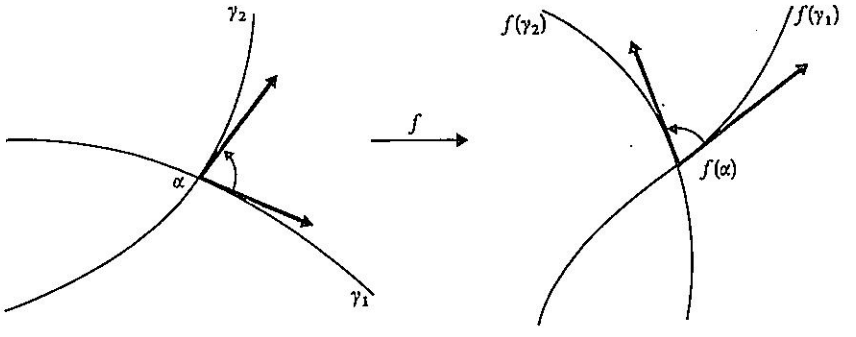

γ_2
α
γ_1
f
$f(\gamma_2)$
$f(\gamma_1)$
$f(\alpha)$

Fig. 2

Proof

We assume without loss of generality that $\gamma_1(0) = \gamma_2(0) = \alpha$.

Since $f'(\alpha) \neq 0$ and f' is continuous, it follows that f' is non-zero on some neighbourhood of α. Now $(f \circ \gamma_1)'(t) = f'(\gamma_1(t)) \cdot \gamma_1'(t)$, so provided t is sufficiently close to 0, we have that $(f \circ \gamma_1)'(t) \neq 0$, and so $f \circ \gamma_1$ is a smooth arc for t in some interval about 0. The same argument applies to $f \circ \gamma_2$.

Note that $(f \circ \gamma_1)(0) = (f \circ \gamma_2)(0) = f(\alpha)$.

The tangent to $f \circ \gamma_1$ at $f(\alpha)$ is $t \longrightarrow f(\alpha) + t \cdot f'(\alpha)\gamma_1'(0)$; a similar expression holds for $f \circ \gamma_2$. The angle between γ_1 and γ_2 at α is

$$\operatorname{Arg} \gamma_2'(0) - \operatorname{Arg} \gamma_1'(0), \quad \text{up to multiples of } 2\pi,$$

and lies between 0 (inclusive) and 2π. The angle between $f \circ \gamma_1$ and $f \circ \gamma_2$ at $f(\alpha)$ is

$$\operatorname{Arg}\left(f'(\alpha) \cdot \gamma_2'(0)\right) - \operatorname{Arg}\left(f'(\alpha) \cdot \gamma_1'(0)\right), \quad \text{up to multiples of } 2\pi,$$

and lies between 0 (inclusive) and 2π. But

$$\begin{aligned}\operatorname{Arg}\left(f'(\alpha) \cdot \gamma_2'(0)\right) &- \operatorname{Arg}\left(f'(\alpha) \cdot \gamma_1'(0)\right)\\ &= \operatorname{Arg} f'(\alpha) + \operatorname{Arg} \gamma_2'(0) - \operatorname{Arg} f'(\alpha) - \operatorname{Arg} \gamma_1'(0)\\ &\quad \text{(up to multiples of } 2\pi)\\ &= \operatorname{Arg} \gamma_2'(0) + \operatorname{Arg} \gamma_1'(0)\\ &\quad \text{(up to multiples of } 2\pi).\end{aligned}$$

Thus the angle between $f \circ \gamma_1$ and $f \circ \gamma_2$ at $f(\alpha)$ differs from the angle between γ_1 and γ_2 at α by a multiple of 2π. But since both angles lie between 0 (inclusive) and 2π, they must be equal. ■

Transformations which have this property of preserving angles between arcs, while not necessarily preserving distance relationships, are called **conformal transformations**. We have shown that a function analytic on a region defines a conformal transformation.

Notice that the function $z \longrightarrow \bar{z}$, which is not analytic, is not, strictly speaking, conformal; although it does not change the magnitudes of angles, it reverses their sense. (To be precise, if the angle between γ_1 and γ_2 at α is θ, then the angle between $\overline{\gamma_1}$ and $\overline{\gamma_2}$ at $\bar{\alpha}$ is $2\pi - \theta$. See Fig. 3.) Since the property of being conformal correlates so conveniently with analyticity, we prefer not to admit conjugation as a conformal transformation.

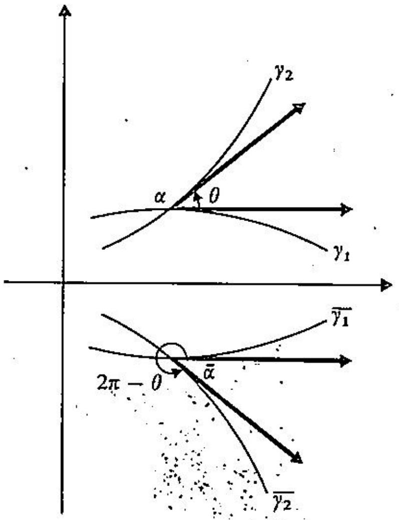


Fig. 3

It can be shown that if the partial derivatives of the real and imaginary parts of a complex function f exist and are continuous on a neighbourhood of α, and if f defines a conformal transformation, then f must be differentiable at α (in the sense of complex analysis), and have non-zero derivative. (See Problem 3, Section 13.2.) Thus being "conformal on a region", and being "analytic and having non-zero derivative on the region", boil down to the same thing.

We shall frequently be interested in transformations which are one-one and analytic on some region. We now show that such a transformation is necessarily conformal.

The proof makes use of some facts about winding numbers. We shall need the same sort of reasoning in the proofs of several theorems in this section, so we shall take this opportunity of reminding you of the facts we shall need.

The winding number of a closed contour about a point is defined as follows. If Γ is the closed contour and α the point (α does not lie on Γ), then the winding number of Γ about α, $\operatorname{Wnd}(\Gamma, \alpha)$, is $\dfrac{1}{2\pi i}\displaystyle\int_\Gamma \frac{dz}{z - \alpha}$. (See *Unit 9, Cauchy's Theorem II*, Section 9.5.)

The main properties of the winding number are these: firstly, $\operatorname{Wnd}(\Gamma, \alpha)$ is always an integer, and in particular if Γ is a simple-closed contour then $\operatorname{Wnd}(\Gamma, \alpha)$ is either ± 1 or 0; secondly, the function $\alpha \longrightarrow \operatorname{Wnd}(\Gamma, \alpha)$ is continuous on the complement of Γ, and thus if R is a region contained in the complement of Γ

then $\mathrm{Wnd}(\Gamma, \alpha)$ is constant for α in R. (These properties are established in Section 9.5 of *Unit 9.*)

We shall also need to use the Principle of the Argument, formulated in terms of winding number. Because the function in question is analytic on a region (and so has no poles), the Principle of the Argument says that if f is a function analytic on a simply-connected region R, Γ is a simple-closed contour contained in R, and f has no zeros on Γ, then $\mathrm{Wnd}(f(\Gamma), 0)$ is the number of zeros of f inside Γ, each zero being counted according to its order. (This last phrase means that a zero of order 3, for example, counts 3 towards the total number of zeros of f inside Γ.) These matters are dealt with in *Unit 10, The Calculus of Residues*, Section 10.8.

We now turn to the theorem about one-one analytic functions.

Theorem 2

If f is analytic and one-one on a region R, then it is conformal.

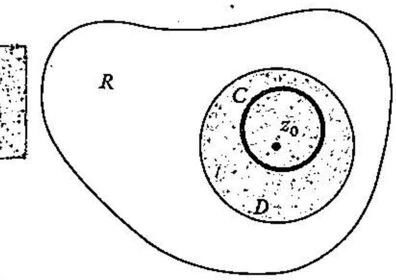


Fig. 4

Proof

We have to show that f' is never zero on R.

Let z_0 be a point of R; let D an open disc contained in R and containing z_0 and C a circle lying in D and surrounding z_0 (Fig. 4). Let $\phi(z) = f(z) - f(z_0)$. Then ϕ is analytic on the simply-connected region D (since it is analytic on the whole of R). It is one-one on D (and on R), and so has but one zero in D, namely z_0. The circle C is a simple-closed contour and ϕ has no zeros on C. Thus by the Principle of the Argument, $\mathrm{Wnd}(\phi(C), 0)$ is the number of zeros of ϕ inside C. Since there is only one zero of ϕ inside C, that at z_0, $\mathrm{Wnd}(\phi(C), 0)$ is in fact the order of this zero. It is at least 1. But since ϕ is one-one, and C is simple-closed, so is $\phi(C)$; therefore $\mathrm{Wnd}(\phi(C), 0)$ is either ± 1 or 0. It must therefore be 1. It follows that z_0 is a simple zero of ϕ, and so $\phi'(z_0) \neq 0$. But $\phi' = f'$, and so $f'(z_0) \neq 0$. Thus f' is never zero on R, and f is conformal. ■

A one-one analytic function defines a conformal transformation. It is natural to wonder whether its inverse is conformal too.

You may remember that we ran into a bit of trouble in Section 3.1 of *Unit 3* when trying to show that the inverse of a differentiable function is differentiable. We are now in a position to clear up that point, and at the same time assert that the inverse of a one-one analytic function *is* conformal.

The problem that arose in *Unit 3* concerned the continuity of the inverse. Now in *Unit 2*, we proved the following result (Theorem 6, page 71):

> Let ϕ be a function continuous at each point of $\mathbf{C}$ and let U be an open subset of $\mathbf{C}$; then $\phi^{-1}(U)$ is open.

(The theorem is true whether or not ϕ is one-one; in the general case, the inverse image $\phi^{-1}(U)$ must be interpreted as the set of points in $\mathbf{C}$ which are mapped to U by ϕ. But if ϕ is one-one then $\phi^{-1}(U)$ can have another meaning, namely, the image of U under the function ϕ^{-1}. Fortunately this ambiguity is only apparent, because the two possible meanings are actually the same.)

Suppose, then that f is one-one and maps $\mathbf{C}$ onto itself. Then if f^{-1} is continuous, the theorem of *Unit 2*, and the fact that $(f^{-1})^{-1} = f$, give this result:

> Let f be a one-one function from $\mathbf{C}$ onto itself, and let U be an open subset of $\mathbf{C}$; then $f(U)$ is open.

In other words, a one-one function whose inverse is continuous has the property that it maps open sets to open sets. A function (not necessarily one-one) which has this property of mapping open sets to open sets is called an **open mapping**.

In the discussion so far we have assumed that the inverse function is continuous, which is just what we need to prove. The point is, that our remarks suggest a method of proving that the inverse function is continuous, namely, to prove firstly that an analytic function is an open mapping, and secondly that the inverse of a one-one open mapping is continuous. This second result is a corollary to the following theorem, in which we prove the first one.

Theorem 3 (The Open Mapping Theorem)

If f is analytic on a region R, and not constant, then it is an open mapping.

Proof

We have to show that if U is an open set contained in R, then $f(U)$ is an open set. In other words, we have to show that if $w_0 \in f(U)$ then there is a neighbourhood of w_0 all of whose points are images of points of U.

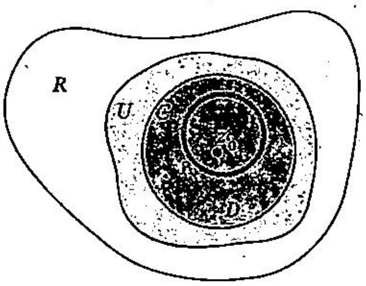


Fig. 5

Since $w_0 \in f(U)$, there is $z_0 \in U$ such that $w_0 = f(z_0)$. Then since U is open there is an open disc D, centre z_0, contained in U. Let $\phi(z) = f(z) - f(z_0)$. Because f is analytic and not constant, it has only isolated zeros (by Corollary 1 to Theorem 15 of *Unit 6, Taylor Series*). So there is a neighbourhood of z_0 which contains no zero of ϕ other than z_0. Thus we may choose a circle C contained in D, surrounding z_0, and having no zeros of ϕ lying on it. (See Fig. 5.) Then $\text{Wnd}(\phi(C), 0)$ is equal to the order of the zero of ϕ inside C, and is consequently at least 1.

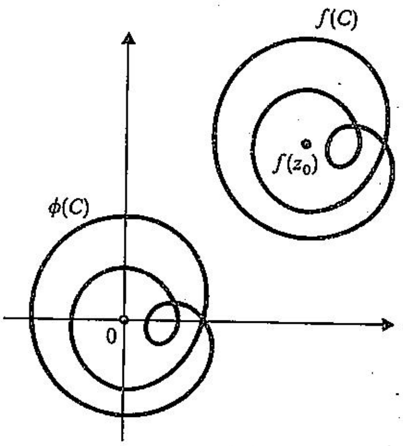


Fig. 6

Now $f(z) = \phi(z) + f(z_0)$, so $f(C)$, the image of the circle C under f, is obtained by translating $\phi(C)$ through $f(z_0)$. (See Fig. 6.) The point 0 is mapped to $f(z_0)$ by this translation. It follows that $\text{Wnd}(\phi(C), 0) = \text{Wnd}(f(C), f(z_0))$. Now let V be the region in the complement of $f(C)$ in which $f(z_0)$ lies (Fig. 7). The function $w \longrightarrow \text{Wnd}(f(C), w)$ is constant on V, and so

$$\begin{aligned}\text{Wnd}(f(C), w) &= \text{Wnd}(f(C), f(z_0))\\ &= \text{Wnd}(\phi(C), 0)\\ &\geqslant 1 \quad \text{for all } w \text{ in } V.\end{aligned}$$

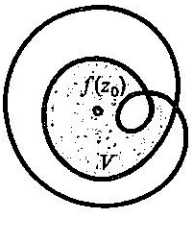


Fig. 7

But $\text{Wnd}(f(C), w)$ is equal to the number of zeros of $z \longrightarrow f(z) - w$ inside C. Thus for any w in V, there is at least one zero of $z \longrightarrow f(z) - w$ inside C, and so at least one z inside C such that $f(z) = w$. The inside of C is contained in U, so we may say that for any w in V, there is at least one point z in U such that $f(z) = w$. Thus $V \subset f(U)$. But V is an open set, in which $w_0 = f(z_0)$ lies. So we have shown that if $w_0 \in f(U)$ then there is a neighbourhood of w_0 contained in $f(U)$. Thus f is an open mapping. ■

Corollary 1

If f is analytic on a region R, and not constant, then $f(R)$ is a region.

Proof

We know from the Open Mapping Theorem that $f(R)$ is open. We have to show that it is connected; this follows from Problem 1 of Section 2.8 of *Unit 2*. Thus $f(R)$ is a region. ■

Corollary 2

If f is analytic, and one-one, on R, then f^{-1} is continuous, and in fact analytic, on $f(R)$, and defines a conformal transformation of $f(R)$ onto R.

Proof

To show that f^{-1} is analytic on $f(R)$ it is enough to show that it is continuous on $f(R)$, since we have already shown (in Theorem 7 of *Unit 3*) that if f^{-1} is continuous at a point w of $f(R)$ then it is differentiable there (provided $f'(f^{-1}(w)) \neq 0$).

The continuity of f^{-1} follows from the Open Mapping Theorem. Let $w_0 \in f(R)$, and let $z_0 = f^{-1}(w_0)$. Let ε be a positive number, and let $U = \{z : |z - z_0| < \varepsilon, z \in R\}$. Then U is an open set, and so $f(U)$ is open. Thus there is $\delta > 0$ such that if $|w - w_0| < \delta$ then $w \in f(U)$. But if $w \in f(U)$, then $f^{-1}(w) \in U$, and so

$$|f^{-1}(w) - z_0| = |f^{-1}(w) - f^{-1}(w_0)| < \varepsilon.$$

So we have shown that if $\varepsilon > 0$, there is $\delta > 0$ such that $|f^{-1}(w) - f^{-1}(w_0)| < \varepsilon$ whenever $|w - w_0| < \delta$; in other words, f^{-1} is continuous at w_0.

Now if $w \in f(R)$, then $(f^{-1})(w) = \dfrac{1}{f'(f^{-1}(w))} \neq 0$. Thus f^{-1} defines a conformal transformation. ■

The situation, then, is as follows: a function which is one-one and analytic on some region defines a mapping of the region which is conformal (angle-preserving) and open (the image of an open set is open); and its inverse has the same properties.

We show next that any conformal transformation is locally one-one. The inclusion of the word "locally" is important: remember that the exponential function is entire, and its derivative is never zero on **C**—but it is not one-one on the whole of **C**. In fact, the proof is an extension of the ideas of the proof of the Open Mapping Theorem. (For a more elementary proof, see Problem 1 of Section 13.2.)

Theorem 4

If f is analytic on a region R, $z_0 \in R$, and $f'(z_0) \neq 0$, then there is a neighbourhood of z_0 on which f is one-one.

Proof

Since $f'(z_0) \neq 0$, there is a disc D contained in R such that f' is never zero on D. The function $\phi(z) = f(z) - f(z_0)$ has a simple zero at z_0, since its derivative at z_0 is not zero. Moreover, we can choose D sufficiently small that it contains no zeros of ϕ apart from z_0, since the zeros of a non-constant analytic function are isolated. Let C be a circle in D surrounding z_0. Then

$$\text{Wnd}(\phi(C), 0) = \text{Wnd}(f(C), f(z_0)) = 1.$$

Let V be the region in the complement of C in which $f(z_0)$ lies. Then the function $w \longrightarrow \text{Wnd}(f(C), w)$ is constant on V and so

$$\text{Wnd}(f(C), w) = \text{Wnd}(f(C), f(z_0)) = 1.$$

So for any w in V, there is exactly one zero of $z \longrightarrow f(z) - w$ inside C, that is, exactly one point z inside C such that $f(z) = w$. That is to say, f is one-one from the inside of C onto V. ■

It follows from Corollary 2 to Theorem 3 that f has an analytic inverse in V, which maps V conformally onto the inside of C.

The question of how one calculates the inverse of f is an interesting one, but is really something of a red herring so far as this section is concerned. We have set two problems in the next problems section in which the matter is dealt with.

We have seen that in a neighbourhood of a point at which its derivative is non-zero, an analytic function is a model of good behaviour. The next question, naturally, is what happens at points at which the derivative of the function vanishes. Consider, for example, the function $z \longrightarrow z^n$. Its derivative at any point other than 0 is non-zero, and so there is a neighbourhood of such a point on which the function is one-one, conformal, and has a conformal inverse. On any punctured neighbourhood of 0, however, the function is "n-one". That is, each point in the range is the image of exactly n points in the punctured neighbourhood of 0. So it certainly cannot have an inverse; and it fails to be conformal at 0—in fact it multiplies angles at 0 by the factor n. This is typical of the general case: a function f for which

$$f'(z_0) = f''(z_0) = \cdots = f^{(n-1)}(z_0) = 0, f^{(n)}(z_0) \neq 0,$$

behaves essentially as the function $z \longrightarrow z^n$ does near 0.

Theorem 5

Suppose that f is analytic on a neighbourhood of z_0, and that $z \longrightarrow f(z) - f(z_0)$ has a zero of order n at z_0. Then there is a one-one analytic function ϕ defined on a neighbourhood U of z_0 such that $f(z) - f(z_0) = (\phi(z))^n$ for z in U. (Thus $z \longrightarrow f(z) - f(z_0)$ is the composition $(z \longrightarrow z^n) \circ \phi$. The function ϕ may be thought of as a "change of coordinates" in the domain of f.)

Proof

Since $z \longrightarrow f(z) - f(z_0)$ has a zero of order n at z_0 there is a function g analytic on a neighbourhood N of z_0 such that

$$f(z) = f(z_0) + (z - z_0)^n g(z), \quad \text{and} \quad g(z_0) \neq 0.$$

Let L be a branch of the logarithm whose domain R contains $g(z_0)$. The inverse image $g^{-1}(N \cap R)$ is an open set containing z_0; let $U' = g^{-1}(N \cap R)$. Then for z in U', the function $z \longrightarrow L(g(z))$ is well defined. Let $\phi(z) = (z - z_0) \exp\left(\frac{1}{n} L(g(z))\right)$. Then for z in U' we have

$$\begin{aligned} (\phi(z))^n &= (z - z_0)^n \exp L(g(z)) \\ &= (z - z_0)^n g(z) \\ &= f(z) - f(z_0). \end{aligned}$$

Moreover, $\phi'(z_0) = \exp\left(\frac{1}{n} L(g(z_0))\right) \neq 0$, so there is a neighbourhood of z_0 on which ϕ is one-one. Take U to be the intersection of U' with this neighbourhood. ■

Summary

This section has been devoted to a study of local mapping properties of analytic functions. We have shown that a function which is analytic on some region and whose derivative is not zero, is conformal, open, locally one-one, and has a conformal local inverse. Near a point at which the derivative of the function is zero the transformation behaves like that defined by $z \longrightarrow z^n$ for some positive integer n.

Self-Assessment Questions

Classify the following statements as True or False.

1. The function exp is everywhere conformal.
2. If the function f satisfies $f'(z) \neq 0$ for z in some region R, then f is one-one on R.
3. Since the function $z \longrightarrow \cos z - 1$ has a zero of order 2 at 0, there is a neighbourhood of 0 on which $\cos z - 1 = z^2$ for all z.
4. The image of the region $\operatorname{Re} z > 1$ under the mapping $z \longrightarrow z^n - 1$ is a region.

Solutions

1. True. (See Theorem 2.)
2. False. (Consider the function exp. The function f is *locally* one-one, but not necessarily one-one on the whole of R.)
3. False. (There is a one-one analytic function ϕ, defined on a neighbourhood of 0, such that $\cos z - 1 = (\phi(z))^2$.)
4. True. (See Corollary 1 to Theorem 3.)

13.2 PROBLEMS

1. If f is analytic on a region containing a point c then for all a and b in some neighbourhood of c we can write

$$f(b) - f(a) = \int_a^b f'(z)dz.$$

(i) If $f'(c) = m \neq 0$, show that there is some neighbourhood D of c such that

$$|f'(z) - f'(c)| < |m| \quad \text{if } z \in D.$$

(ii) If $a, b \in D$ and $f(a) = f(b)$, show that

$$(b - a)f'(c) + \int_a^b [f'(z) - f'(c)]dz = 0$$

and deduce that $a = b$.

2. Let $f = U + iV$ be a function defined on a region R, with real and imaginary parts U and V respectively. Suppose that U and V have continuous partial derivatives on R. Let γ_α be the image under f of the straight line $t \longrightarrow z_0 + te^{i\alpha}$ (the line through z_0, a point of R, which makes angle α with the real axis). Show that

$$\gamma_\alpha'(0) = \frac{\partial U}{\partial x}(x_0, y_0) \cdot \cos\alpha + \frac{\partial U}{\partial y}(x_0, y_0) \cdot \sin\alpha$$
$$+ i\left(\frac{\partial V}{\partial x}(x_0, y_0) \cdot \cos\alpha + \frac{\partial V}{\partial y}(x_0, y_0) \cdot \sin\alpha\right).$$

(You may need to use the formula for the directional derivative of a function of two variables, ϕ, which is

$$\nabla_\alpha \phi(x_0, y_0) = \frac{\partial \phi}{\partial x}(x_0, y_0) \cdot \cos\alpha + \frac{\partial \phi}{\partial y}(x_0, y_0) \cdot \sin\alpha.$$

This is given in *Unit M231 5, Differentiation*, Section 5.9, Theorem 4.)

Suppose that f maps smooth curves to smooth curves, and is conformal—that is, preserves angles. Show that f is analytic, and has non-zero derivative.

(Hint: Aim to establish the Cauchy–Riemann equations. The special case in which the tangent to γ_0 is parallel to the real axis is straightforward. Can you reduce the general case to this special case?)

3. Deduce the Local Maximum Principle from the Open Mapping Theorem. That is, show that if f is analytic on a region R, and not constant on R, then $|f|$ has no local maximum at any point in R. (This is Theorem 1, of *Unit 9*.)

(Your argument should work for any open mapping, even one which is not defined by a non-constant analytic function. One could say that the Local Maximum Principle is really a theorem of topology, not of complex analysis at all.)

The remaining two problems, both of which are rather long, deal with the question of finding the local inverse of an analytic function f whose derivative at some point z_0 is non-zero.

The first method, outlined in Problem 4, is an iterative method due to Picard. It is a method commonly used in such situations, so you may have come across it before. For example, it is very similar to the method used to prove the existence of solutions of ordinary differential equations in *Unit M201 33, Existence and Uniqueness Theorem for Differential Equations.*

In Problem 5 we derive a formula for the Taylor series of the inverse function. This formula is due to Lagrange.

You should attempt only one of these problems unless you are feeling particularly energetic, but you should read both solutions.

---

4. *Picard's method for finding local inverses*
We assume, for simplicity, that we are dealing with a function f such that $f(0) = 0$, and $f'(0) = 1$. The general case can be treated by appropriate modifications of the method. Thus we have $f(z) = z + \phi(z)$, where ϕ is analytic, and $\phi(0) = \phi'(0) = 0$. We wish to solve the equation $f(z) = w$ for z, given w sufficiently close to 0; we define $f^{-1}(w)$ to be the solution of this equation, which we know exists and is unique. The equation may be written

$$z = w - \phi(z),$$

which suggests an iterative scheme

$$z_{n+1} = w - \phi(z_n),$$

with $z_0 = 0$ say. The task, then, is to show that the sequence $\{z_n\}$, thus defined, converges to a solution of the original equation.

(a) Show that if the sequence does converge, to the limit z say, then $z = w - \phi(z)$ as required.

(b) Since $\phi'(0) = 0$, there is a disc $D = \{z: |z| < r\}$ such that $|\phi'(z)| < K$, where K is some fixed number less than 1, for all z in D. By considering the equation

$$\phi(\alpha) - \phi(\beta) = \int_{[\alpha,\beta]} \phi'(\zeta)d\zeta,$$

show that $|\phi(\alpha) - \phi(\beta)| < K|\alpha - \beta|$ for all α and β in D.

(c) Show that if $|w| < (1 - K)r$, then $z_n \in D$ for every $n \geqslant 0$.

(d) Deduce that for every $n \geqslant 1$,

$$|z_{n+1} - z_n| < K|z_n - z_{n-1}|$$

and hence that $\{z_n\}$ is a Cauchy sequence.

(e) Conclude that $\{z_n\}$ converges.

(f) Calculate a few terms (at least five) in the (local) inverse of $z \longrightarrow z + z^2$ by this method.

5. *Lagrange's formula for the Taylor series of the inverse*
Suppose that $f(0) = 0$ and $f'(0) \neq 0$, for simplicity's sake. The method can be adapted to the general case. We know that f has an inverse f^{-1} which is analytic on a neighbourhood of 0, and so may be expanded in a Taylor series about 0: suppose

$$f(z) = a_1z + a_2z^2 + a_3z^3 + \cdots$$

and

$$f^{-1}(w) = b_1w + b_2w^2 + b_3w^3 + \cdots.$$

(There are no constant terms since $f(0) = 0$ and $f^{-1}(0) = 0$.)

(a) By substituting in the formula $f^{-1}(f(z)) = z$, calculate a few of the coefficients $b_1, b_2, \ldots$ in terms of $a_1, a_2, \ldots$. This problem should convince you that this straightforward way of finding the Taylor series of f^{-1} will never give a simple formula for all b_n in terms of f. We have to do something more sophisticated.

(b) Denote f^{-1} by F. Then, we have

$$b_n = \frac{F^{(n)}(0)}{n!}, \quad \text{by Taylor's Theorem,}$$
$$= \frac{1}{2\pi i}\int_C \frac{F(w)}{w^{n+1}}dw,$$

by Cauchy's Formula for nth derivative, C being a suitable circle.

Let $z = F(w)$; use this substitution and then integration by parts to show that

$$b_n = \frac{1}{n}\operatorname{Res}(z \longrightarrow 1/[f(z)]^n, 0),$$

and hence that

$$f^{-1}(w) = \sum_{n=1}^{\infty} \frac{1}{n!} \lim_{z\to 0} \left[\frac{d^{n-1}}{dz^{n-1}} \left(\frac{z}{f(z)} \right)^n \right] \cdot w^n.$$

(c) Calculate a few terms (at least five) in the Taylor series of the inverse of $z \longrightarrow z + z^2$, using the formula of part (b).

Solutions

1. (i) Since f is differentiable at c it follows that f' is also differentiable, and therefore continuous, at c. The required result follows from the definition of continuity.

(ii) We have

$$\begin{aligned}(b-a)f'(c) + \int_a^b [f'(z) - f'(c)]\,dz &= \int_a^b f'(z)\,dz \\ &= f(b) - f(a) = 0,\end{aligned}$$

and therefore

$$\begin{aligned}|(b-a)f'(c)| &= \left| \int_a^b [f'(z) - f'(c)]\,dz \right| \\ &\leqslant |b-a| \cdot |m|.\end{aligned}$$

But $|f'(c)| = |m|$ and therefore $|b - a| = 0$, that is $a = b$. (This is an alternative proof of Theorem 4.)

2. We have

$$\begin{aligned}\gamma_\alpha(t) &= f(z_0 + te^{i\alpha}) \\ &= U(x_0 + t\cos\alpha, y_0 + t\sin\alpha) + iV(x_0 + t\cos\alpha, y_0 + t\sin\alpha).\end{aligned}$$

Thus

$$\begin{aligned}\gamma_\alpha'(0) &= \nabla_\alpha U(x_0, y_0) + i\nabla_\alpha V(x_0, y_0) \\ &= \frac{\partial U}{\partial x}(x_0, y_0)\cdot\cos\alpha + \frac{\partial U}{\partial y}(x_0, y_0)\cdot\sin\alpha \\ &\quad + i\left(\frac{\partial V}{\partial x}(x_0, y_0)\cdot\cos\alpha + \frac{\partial V}{\partial y}(x_0, y_0)\cdot\sin\alpha\right).\end{aligned}$$

(The functions U and V are differentiable at (x_0, y_0) since they have continuous partial derivatives.)

Suppose that f is conformal, and that the tangent to γ_0 at $f(z_0)$ is parallel to the real axis. Then

$$\gamma_0'(0) = \frac{\partial U}{\partial x}(x_0, y_0) + i\frac{\partial V}{\partial V}(x_0, y_0)$$

is real, and so

$$\frac{\partial V}{\partial x}(x_0, y_0) = 0.$$

Now since f is conformal, it must be true that the tangent to γ_α at $f(z_0)$ makes angle α with the real axis. In particular, the tangent to $\gamma_{\pi/2}$ is parallel to the imaginary axis; but

$$\gamma_{\pi/2}'(0) = \frac{\partial U}{\partial y}(x_0, y_0) + i\frac{\partial V}{\partial y}(x_0, y_0),$$

and so

$$\frac{\partial U}{\partial y}(x_0, y_0) = 0.$$

Also

$$\gamma'_{\pi/4}(0) = \frac{1}{\sqrt{2}}\left(\frac{\partial U}{\partial x}(x_0, y_0) + \frac{\partial U}{\partial y}(x_0, y_0)\right) + \frac{i}{\sqrt{2}}\left(\frac{\partial V}{\partial x}(x_0, y_0) + \frac{\partial V}{\partial y}(x_0, y_0)\right)$$
$$= \frac{1}{\sqrt{2}}\left(\frac{\partial U}{\partial x}(x_0, y_0) + i\frac{\partial V}{\partial y}(x_0, y_0)\right).$$

Since the tangent to $\gamma_{\pi/4}$ at $f(z_0)$ must make angle $\pi/4$ with the real axis, we must have

$$\frac{\partial U}{\partial x}(x_0, y_0) = \frac{\partial V}{\partial y}(x_0, y_0).$$

Thus the Cauchy–Riemann equations are satisfied (the equation $\frac{\partial U}{\partial y}(x_0, y_0) = -\frac{\partial V}{\partial x}(x_0, y_0)$ being satisfied since both sides are zero). This means that f is differentiable at z_0. Moreover, since f maps smooth curves to smooth curves, the partial derivatives $\frac{\partial U}{\partial x}(x_0, y_0)$, $\frac{\partial U}{\partial y}(x_0, y_0)$, $\frac{\partial V}{\partial x}(x_0, y_0)$ and $\frac{\partial V}{\partial y}(x_0, y_0)$ cannot all be zero. Thus the derivative of f, which is $\frac{\partial U}{\partial x}(x_0, y_0)$, is nonzero.

Now consider a function f for which the tangent to γ_0 at $f(z_0)$ is not parallel to the real axis. Suppose that this tangent makes angle θ with the real axis, so that $\operatorname{Arg} \gamma'_0(0) = \theta$. If f is followed by a rotation about $f(z_0)$ through angle $-\theta$, then the resulting composite function will map the line through z_0 parallel to the real axis to a curve whose tangent at $f(z_0)$ *is* parallel to the real axis. It will also be conformal. The composite function is g, where

$$g(z) = f(z_0) + e^{-i\theta}(f(z) - f(z_0))$$

Let $f = U + iV$. Then

$$g(z) = (1 - e^{-i\theta})(U(x_0, y_0) + iV(x_0, y_0))$$
$$+(\cos\theta \cdot U + \sin\theta \cdot V) + i(\cos\theta \cdot V - \sin\theta \cdot U).$$

The partial derivatives of the real and imaginary parts of g satisfy the Cauchy–Riemann equations. Thus

$$\cos\theta\frac{\partial U}{\partial x}(x_0, y_0) + \sin\theta\frac{\partial V}{\partial x}(x_0, y_0) = \cos\theta\frac{\partial V}{\partial y}(x_0, y_0) - \sin\theta\frac{\partial U}{\partial y}(x_0, y_0)$$

and

$$\cos\theta\frac{\partial U}{\partial y}(x_0, y_0) + \sin\theta\frac{\partial V}{\partial y}(x_0, y_0) = \sin\theta\frac{\partial U}{\partial x}(x_0, y_0) - \cos\theta\frac{\partial V}{\partial x}(x_0, y_0).$$

If we multiply the first equation by $\cos\theta$ and the second by $-\sin\theta$ and add, we obtain

$$\frac{\partial U}{\partial x}(x_0, y_0) = \frac{\partial V}{\partial y}(x_0, y_0).$$

If we multiply the first equation by $\sin\theta$ and the second by $-\cos\theta$ and add, we obtain

$$\frac{\partial V}{\partial x}(x_0, y_0) = -\frac{\partial U}{\partial y}(x_0, y_0).$$

Thus f is differentiable at z_0. Since f maps smooth curves to smooth curves its derivative cannot be zero.

3. Suppose that f is a complex function (not necessarily analytic) whose domain is a region R. Suppose that f is an open mapping: thus if U is an open subset of R, then $f(U)$ is open. We have to show that $|f|$ has no local maximum at any point of R.

   Suppose that $|f|$ did have a local maximum at some point z_0 of R. Then there would be some $\delta > 0$ such that $|f(z)| \leqslant |f(z_0)|$ for all z such that $|z - z_0| < \delta$. (We may of course assume that δ is sufficiently small that all such z lie in R.) But if w_0 is an interior point of a set V, then there is necessarily some point w in V such that $|w| > |w_0|$ (Fig. 8).

   Thus if $|f|$ did have a local maximum at z_0, then $f(z_0)$ could not be an interior point of the image of the open disc $\{z: |z - z_0| < \delta\}$. That image could not be open. But this would contradict the assumption that f is an open mapping. This establishes the Local Maximum Principle for any open mapping. In particular, a non-constant function, analytic on a region R, is an open mapping, and so obeys the Local Maximum Principle.

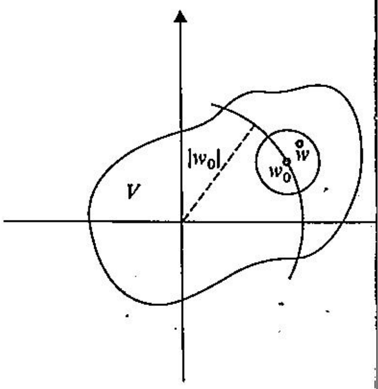


Fig. 8

4. (a) If $\lim_{n\to\infty} z_n = z$, then since ϕ is continuous, $\lim_{n\to\infty} \phi(z_n) = \phi(z)$. Thus

$$z = \lim_{n\to\infty} z_{n+1} = \lim_{n\to\infty} (w - \phi(z_n)) = w - \phi(z).$$

(b) By the Estimation Theorem

$$\begin{aligned}|\phi(\alpha) - \phi(\beta)| &\leqslant \int_{[\alpha,\beta]} |\phi'(\zeta)| \cdot |d\zeta| \\ &< K|\alpha - \beta|.\end{aligned}$$

(c) We prove this by mathematical induction. Clearly $z_0 \in D$. Suppose that $z_n \in D$. Then

$$\begin{aligned}|z_{n+1}| &= |w - \phi(z_n)| \\ &\leqslant |w| + |\phi(z_n)| \\ &< (1 - K)r + K|z_n|, \quad \text{using part (b),} \\ &< (1 - K)r + Kr = r.\end{aligned}$$

Thus $z_{n+1} \in D$.

(d)
$$\begin{aligned}|z_{n+1} - z_n| &= |w - \phi(z_n) - (w - \phi(z_{n-1}))| \\ &= |\phi(z_{n-1}) - \phi(z_n)| \\ &< K|z_n - z_{n-1}|, \quad \text{since } z_{n-1} \text{ and } z_n \text{ lie in } D. \qquad (1)\end{aligned}$$

Using this result repeatedly, we have for $k \geqslant 1$,

$$\begin{aligned}|z_{n+k} - z_{n+k-1}| &< K|z_{n+k-1} - z_{n+k-2}| \\ &< K^2|z_{n+k-2} - z_{n+k-3}| \\ &\dots \\ &< K^{k-1}|z_{n+1} - z_n|. \qquad (2)\end{aligned}$$

Thus

$$\begin{aligned}|z_{n+p} - z_n| &\leqslant |z_{n+p} - z_{n+p-1}| + |z_{n+p-1} - z_{n+p-2}| + \cdots + |z_{n+1} - z_n| \\ &< K^{p-1}|z_{n+1} - z_n| + K^{p-2}|z_{n+1} - z_n| + \cdots + |z_{n+1} - z_n|, \quad \text{by (2),} \\ &< (K^{p-1} + K^{p-2} + \cdots + 1)|z_{n+1} - z_n| \\ &= \frac{1 - K^p}{1 - K}|z_{n+1} - z_n| \\ &< \frac{1}{1 - K}|z_{n+1} - z_n| \\ &< \frac{K^n}{1 - K}|z_1 - z_0|, \quad \text{by repeated use of (1),} \\ &= \frac{K^n}{1 - K}|w|, \quad \text{since we choose } z_0 = 0, \text{ whence } z_1 = w.\end{aligned}$$

Now $0 < K < 1$, and so $\lim_{n\to\infty} K^n = 0$. It follows that $\{z_n\}$ is a Cauchy sequence.

(e) Since $\{z_n\}$ is a Cauchy sequence, it converges to some point z, such that $z = w - \phi(z)$, that is $f(z) = w$.

(f) If $f(z) = z + z^2$, then $f(0) = 0, f'(0) = 1$, and $\phi(z) = z^2$. Then $|\phi'(z)| < \frac{1}{2}$ provided $|z| < \frac{1}{4}$, so we can take $K = \frac{1}{2}$, $r = \frac{1}{4}$. The method then applies for those w such that $|w| < \frac{1}{8}$. We have

$$\begin{aligned}z_0 &= 0; \\ z_1 &= w - z_0^2 = w; \\ z_2 &= w - z_1^2 = w - w^2; \\ z_3 &= w - (w - w^2)^2 = w - w^2 + 2w^3 - w^4; \\ z_4 &= w - (w - w^2 + 2w^3 - w^4)^2 \\ &= w - w^2 + 2w^3 - 5w^4 + 6w^5 - 6w^6 + 4w^7 - w^8.\end{aligned}$$

(These are not the first eight terms of the Taylor series, as you will see in Problem 5.)

5. (a) We have

$$\begin{aligned}f^{-1}(f(z)) &= z\\ &= b_1(a_1z + a_2z^2 + a_3z^3 + \cdots) + b_2(a_1z + a_2z^2 + a_3z^3 + \cdots)^2\\ &\quad + b_3(a_1z + a_2z^2 + a_2z^3 + \cdots)^3 + \cdots\\ &= a_1b_1z + (a_2b_1 + a_1^2\, b_2)z^2 + (a_3b_1 + 2a_1a_2b_2 + a_1^3b_3)z^3 + \cdots\end{aligned}$$

Thus

$$\begin{aligned}a_1b_1 &= 1\\ a_2b_1 + a_1^2b_2 &= 0\\ a_3b_1 + 2a_1a_2b_2 + a_1^3b_3 &= 0,\end{aligned}$$

and so on,

whence

$$\begin{aligned}b_1 &= \frac{1}{a_1}\\ b_2 &= -\frac{a_2}{a_1^3}\\ b_3 &= 2\frac{a_2^2}{a_1^5} - \frac{a_3}{a_1^4}.\end{aligned}$$

Although in principle we could calculate any coefficient b_n simply by writing down enough equations obtained by equating coefficients of z^k to zero, it is not at all clear what the general form of b_n will be.

(b) Let $z = F(w)$; then $w = f(z)$, since $F = f^{-1}$. Making this substitution in

$$b_n = \frac{1}{2\pi i}\int_C \frac{F(w)}{w^{n+1}}dw,$$

we obtain

$$\begin{aligned}b_n &= \frac{1}{2\pi i}\int_C \frac{zf'(z)}{[f(z)]^{n+1}}dz\\ &= \left.\frac{z}{2\pi in[f(z)]^n}\right|_C + \frac{1}{2\pi in}\int_C \frac{dz}{[f(z)]^n}, \quad \text{by integration by parts.}\end{aligned}$$

The only singularities of $z \longrightarrow 1/[f(z)]^n$ are poles at zero and so the first term is zero. The second term is simply $\frac{1}{n}\text{Res}(z \longrightarrow 1/[f(z)]^n, 0)$, as required.

From page 17 of *Unit 10*, we know that if g has a pole of order n at 0 then

$$\text{Res}(g, 0) = \frac{1}{(n-1)!}\lim_{z\to 0}\left[\frac{d^{n-1}}{dz^{n-1}}(g(z)z^n)\right].$$

Using this result, we have

$$b_n = \frac{1}{n}\cdot\frac{1}{(n-1)!}\lim_{z\to 0}\left[\frac{d^{n-1}}{dz^{n-1}}\left(\frac{z}{f(z)}\right)^n\right],$$

and so

$$F(w) = f^{-1}(w) = \sum_{n=1}^{\infty}\frac{1}{n!}\lim_{z\to 0}\left[\frac{d^{n-1}}{dz^{n-1}}\left(\frac{z}{f(z)}\right)^n\right]\cdot w^n.$$

(c) We have $\frac{z}{f(z)} = \frac{1}{1+z}$. We shall denote

$$\lim_{z\to 0}\left[\frac{d^{n-1}}{dz^{n-1}}\left(\frac{z}{f(z)}\right)^n\right] \quad \text{by} \quad \left[\frac{d^{n-1}}{dz^{n-1}}\left(\frac{z}{f(z)}\right)^n\right]_{z=0}.$$

Thus

$$\left[\frac{d}{dz}\left(\frac{z}{f(z)}\right)^2\right]_{z=0} = \left[\frac{d}{dz}\left(\frac{1}{1+z}\right)^2\right]_{z=0} = -2;$$

$$\left[\frac{d^2}{dz^2}\left(\frac{z}{f(z)}\right)^3\right]_{z=0} = \left[\frac{d^2}{dz^2}\left(\frac{1}{1+z}\right)^3\right]_{z=0} = 12;$$

$$\left[\frac{d^3}{dz^3}\left(\frac{z}{f(z)}\right)^4\right]_{z=0} = \left[\frac{d^3}{dz^3}\left(\frac{1}{1+z}\right)^4\right]_{z=0} = -120;$$

$$\left[\frac{d^4}{dz^4}\left(\frac{z}{f(z)}\right)^5\right]_{z=0} = \left[\frac{d^4}{dz^4}\left(\frac{1}{1+z}\right)^5\right]_{z=0} = 1680.$$

So the first five terms in the Taylor series for $f^{-1}(w)$ are

$$w - w^2 + 2w^3 - 5w^4 + 14w^5.$$

13.3 BILINEAR TRANSFORMATIONS

In this section we shall discuss the geometric properties of a particular kind of transformation, the bilinear transformation.*

* * * * * * * *

Definition

A transformation of the form $z \longrightarrow \dfrac{az + b}{cz + d}$, where a, b, c and d are complex numbers such that $ad - bc \neq 0$, is called a **bilinear transformation.**

In Problem 1 of Section 13.4 we ask you to establish some of the following properties of bilinear transformations.

A bilinear transformation with $c \neq 0$ is a conformal analytic transformation of the region $\mathbf{C} - \{-d/c\}$ onto the region $\mathbf{C} - \{a/c\}$; if $c = 0$ then the bilinear transformation maps the whole complex plane onto itself, and is entire. Notice that if $c \neq 0$ then the function $z \longrightarrow \dfrac{az + b}{cz + d}$ has a singularity at $-d/c$, which is a simple pole.

A bilinear transformation is one-one. For suppose that

$$\frac{az_1 + b}{cz_1 + d} = \frac{az_2 + b}{cz_2 + d};$$

then

$$0 = \frac{az_1 + b}{cz_1 + d} - \frac{az_2 + b}{cz_2 + d} = \frac{(az_1 + b)(cz_2 + d) - (az_2 + b)(cz_1 + d)}{(cz_1 + d)(cz_2 + d)}$$

$$= \frac{(ad - bc)(z_1 - z_2)}{(cz_1 + d)(cz_2 + d)}.$$

Thus $z_1 = z_2$, since $ad - bc \neq 0$, and so the transformation is one-one. (This result explains the significance of the condition $ad - bc \neq 0$. In fact if $ad - bc = 0$, then $z \longrightarrow \dfrac{az + b}{cz + d}$ is a constant function, if it is defined at all.) The inverse of the bilinear transformation $z \longrightarrow \dfrac{az + b}{cz + d}$ is the transformation $z \longrightarrow \dfrac{dz - b}{-cz + a}$, which is again a bilinear transformation. It is easy to check that the composition of two bilinear transformations is again a bilinear transformation.

The Basic Bilinear Transformations

By imposing certain conditions on the numbers a, b, c and d we pick out various types of bilinear transformation. For example, a transformation $z \longrightarrow z + b$ is a bilinear transformation for which $a = 1$, $c = 0$ and $d = 1$. It is a **translation** (see *Unit 2*, page 57). If we take $b = c = 0$, $d = 1$, we obtain a transformation $z \longrightarrow az$. If $|a| = 1$ this is a **rotation about the origin** (see *Unit 2*, page 57); if a is a real, positive number it is a **magnification** (strictly speaking, a magnification if $a > 1$ and a contraction if $a < 1$). These are illustrated in Figs. 9, 10 and 11. Finally, the bilinear transformation $z \longrightarrow 1/z$, for which $a = 0$, $b = c = 1$, and $d = 0$, is **inversion in the unit circle**. If $w = 1/z$, then $|w| = 1/|z|$ and $\operatorname{Arg} w = -\operatorname{Arg} z$; so to find the image of the point z under this mapping, we first find the point z' on the line joining z to the origin such that $|z'| \cdot |z| = 1$, and then take the reflection of z' in the real axis (Fig. 12).

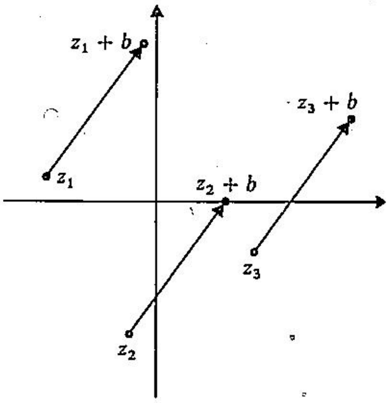


Fig. 9

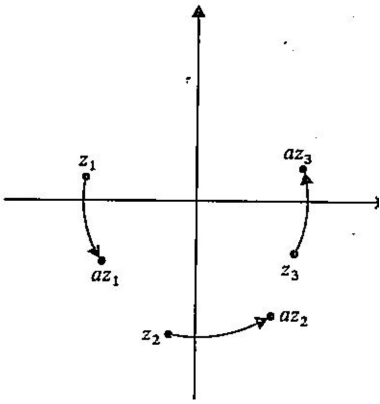


Fig. 10

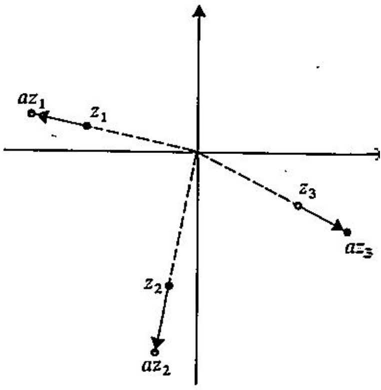


Fig. 11

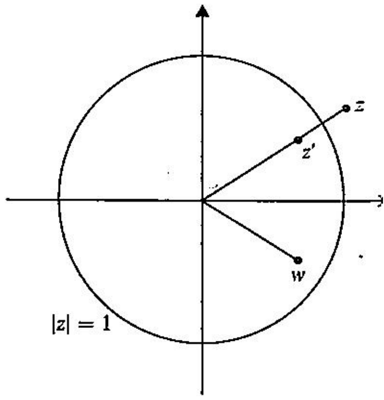


Fig. 12

---

* Any student of M201 will see that a bilinear transformation is not bilinear in the sense that a bilinear form is, but it is standard terminology.

The point of picking out these particular types of bilinear transformation is that they are the basic ones. Any bilinear transformation can be expressed as a composition of transformations of these basic types. To show that this is so we have to consider two cases. Firstly, if $c = 0$, then $d \neq 0$ since $ad - bc \neq 0$. Then $\frac{az+b}{cz+d} = \frac{a}{d}z + \frac{b}{d}$, and $z \longrightarrow \frac{a}{d}z + \frac{b}{d}$ is the composition of a rotation about the origin, a magnification and a translation. If $c \neq 0$, on the other hand, then

$$\frac{az+b}{cz+d} = \frac{\frac{a}{c}(cz+d) + b - \frac{ad}{c}}{cz+d}$$

$$= \frac{a}{c} - \frac{ad-bc}{c}\cdot\frac{1}{cz+d}.$$

This is the composition $h \circ g \circ f$ of the transformations

$$f\colon z \longrightarrow cz + d,$$

$$g\colon z \longrightarrow \frac{1}{z},$$

and

$$h\colon z \longrightarrow -\frac{ad-bc}{c}z + \frac{a}{c}.$$

Clearly the first and last of these can be further broken down into a rotation about the origin, a magnification and a translation.

The basic types of bilinear transformation are all of importance in geometry. This is one of the reasons why, in this unit, we are interested in bilinear transformations.

Firstly, translations and rotations about the origin are examples of isometries. An **isometry** is a function f such that $|f(z_1) - f(z_2)| = |z_1 - z_2|$ for every pair of points z_1 and z_2. It is thus a transformation of the plane which preserves distance relationships. The notion of congruence between figures is very much tied up with the notion of isometry. One could adopt the following definition of congruence, for triangles, say: two triangles whose vertices are z_1, z_2, z_3 and w_1, w_2, w_3 are congruent if there is an isometry f such that $w_1 = f(z_1)$, $w_2 = f(z_2)$ and $w_3 = f(z_3)$.

You may recollect that isometries were the subject of two of the problems in *Unit 2*. In one of these problems (Problem 3, Section 2.2) you were asked to show that the composition of two isometries is an isometry; in other words, that if f and g are isometries so is $f \circ g$. The composition of a rotation about the origin, $z \longrightarrow az$ where $|a| = 1$, followed by a translation $z \longrightarrow z + b$ gives the isometry $z \longrightarrow az + b$. In fact, this isometry is a rotation about some point other than the origin, if $a \neq 1$. To show that this is so we shall find the function which corresponds to a rotation about a point z_0. This rotation has the same effect as first carrying out a translation which takes z_0 to the origin, then rotating about the origin, then translating the origin back to z_0. It is therefore the composition $h \circ g \circ f$ of the functions

$$f\colon z \longrightarrow z - z_0,$$

$$g\colon z \longrightarrow az, \quad \text{where } |a| = 1,$$

$$h\colon z \longrightarrow z + z_0,$$

which is $z \longrightarrow a(z - z_0) + z_0$. If you compare this transformation with $z \longrightarrow az + b$ you will see that the latter is a rotation about $b/(1-a)$, provided $a \neq 1$. When $a = 1$ it is of course a translation.

It can be shown that an isometry is necessarily of one of the following forms: either $z \longrightarrow az + b$, where $|a| = 1$; or $z \longrightarrow a\bar{z} + b$, where $|a| = 1$. (If you need to be convinced that this is so, try Problem 8 of Section 13.4.) The conjugation function $z \longrightarrow \bar{z}$ defines the transformation of **reflection in the real axis**. Thus an isometry is the composition of possibly a reflection, a rotation about the origin, and a translation. But notice that reflection is not an analytic transformation; nor is it a bilinear transformation.

The functions $f(z) = az + b$, where a is non-zero but not necessarily of modulus one, are called **similarities**. They are compositions of magnifications, rotations about the origin, and translations. They satisfy the condition $|f(z_1) - f(z_2)| = |a| \cdot |z_1 - z_2|$ for every pair of points z_1 and z_2; so they do not preserve distances, but they do preserve the ratios of distances. In the same way that isometries are associated with the geometrical notion of congruence, similarities are associated with the geometrical notion of similarity (as in similar triangles).

Finally, the transformation of inversion, $z \longrightarrow 1/z$, is closely related to a transformation which is sometimes useful in geometry. If C is a circle, centre α and radius r, then two points z_1 and z_2 are said to be **inverse with respect to** C if α, z_1 and z_2 are collinear, z_1 and z_2 lie on the same side of α and

$$|z_1 - \alpha| \cdot |z_2 - \alpha| = r^2.$$

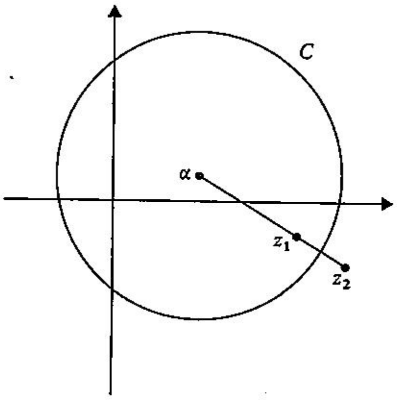


Fig. 13

In particular, the points z and $1/\bar{z}$ (where $z \neq 0$) are inverse with respect to the unit circle. The transformation of geometric inversion with respect to the unit circle is the one in which each point is mapped to its inverse, and so is given by the function $z \longrightarrow 1/\bar{z}$. But this is not an analytic transformation. Inversion is geometric inversion followed by reflection in the real axis.

We shall be in a better position to see why inversion is a useful geometric transformation once we have established some of the properties of bilinear transformations (which are, in particular, properties of inversion).

Images of Circles and Straight Lines

One important property of bilinear transformations is that they map circles and straight lines to circles and straight lines. In other words, the image of a circle under a bilinear transformation is either a circle or a straight line, and the image of a straight line is either a circle or a straight line, We shall prove this shortly; but first we must establish some preliminary results about circles and straight lines.

As you know a circle may be specified by an equation of the form $|z - z_0| = r$. The point z_0 is the centre of the circle, and r is its radius; r is a real, positive number.

Now if $|z - z_0| = r$, then

$$(z - z_0)(\bar{z} - \overline{z_0}) = r^2,$$

that is

$$z\bar{z} - \overline{z_0}z - z_0\bar{z} + (z_0\overline{z_0} - r^2) = 0.$$

Conversely, if

$$z\bar{z} - \overline{z_0}z - z_0\bar{z} + (z_0\overline{z_0} - r^2) = 0,$$

then

$$|z - z_0|^2 = r^2,$$

and so, since r is positive,

$$|z - z_0| = r.$$

Thus the circle $|z - z_0| = r$ is specified equally well by the equation

$$z\bar{z} - \overline{z_0}z - z_0\bar{z} + (z_0\overline{z_0} - r^2) = 0. \qquad (1)$$

(Note that what we have established is the equivalence of the sets $\{z : |z - z_0| = r\}$ and $\{z : z\bar{z} - \overline{z_0}z - z_0\bar{z} + (z_0\overline{z_0} - r^2) = 0\}$.)

Consider, now, the equation

$$Az\bar{z} + \bar{B}z + B\bar{z} + C = 0,$$

where A and C are real numbers, and B is a complex number. If $A \neq 0$, we may write this equation as

$$z\bar{z} + \frac{\bar{B}}{A}z + \frac{B}{A}\bar{z} + \frac{C}{A} = 0;$$

comparing this with (1), we see that this will be the equation of a circle centre $-\dfrac{B}{A}$, provided that

$$\frac{B\bar{B}}{A^2} - \frac{C}{A} > 0. \qquad (2)$$

We require this *validity condition* to hold in order that the radius of the circle, which is $\dfrac{1}{|A|}\sqrt{|B|^2 - AC}$, be well-defined.

If $A = 0$, on the other hand, we are left with the equation

$$\bar{B}z + B\bar{z} + C = 0.$$

It is easy to show that this is the equation of a straight line, and that any straight line can be defined by such an equation. (One way of doing this is to set $z = x + iy$ and use the familiar equation of a straight line from coordinate geometry.)

There is an advantage, in what follows, in using the equation

$$Az\bar{z} + \bar{B}z + B\bar{z} + C = 0,$$

since it allows us to discuss circles and straight lines at the same time. Notice that if $C = 0$, then the curve defined by this equation passes through the origin.

Self-Assessment Questions

1. Which of the following functions define bilinear transformations?

   (i) $z \longrightarrow z + 2i$;

   (ii) $z \longrightarrow \dfrac{z + i}{iz - 1}$;

   (iii) $z \longrightarrow \dfrac{z + i}{z - 1}$;

   (iv) $z \longrightarrow z + \dfrac{1}{z}$.

The Library
Exeter College
Oxford

2. For the following sets of coefficients, state whether the equation $Az\bar{z} + \bar{B}z + B\bar{z} + C = 0$ represents a circle, a straight line, neither a circle nor a straight line.

   (i) $A = C = i, B = 1$;

   (ii) $A = C = 1, B = i$;

   (iii) $A = 1, B = i, C = 0$.

Solutions

1. The functions numbered (i) and (iii) define bilinear transformations. Number (ii) has $ad - bc = 0$, and is in fact the constant function $z \longrightarrow -i$, and number (iv) is $z \longrightarrow \dfrac{z^2 + 1}{z}$.

2. (i) Neither: A and C are not real.

   (ii) Neither: the validity condition (2) is not satisfied.

   (iii) Circle.

To show that, under a bilinear transformation, a circle or straight line is mapped to a circle or straight line, it will be enough to show that the same is true for each of the basic bilinear transformations. We have seen that any bilinear transformation may be expressed as the composition of translations, rotations, magnifications, and inversions. Thus if each of these transformations maps circles and straight lines to circles and straight lines, so will an arbitrary bilinear transformation.

Now it is pretty obvious intuitively that translation, rotation and magnification map circles to circles and straight lines to straight lines. We can show algebraically that this is true as follows. Translation, rotation and magnification are all particular cases of transformations of the form $z \longrightarrow az + b$, where $a \neq 0$ (in other words, of bilinear transformations $z \longrightarrow \dfrac{az + b}{cz + d}$ with $c = 0$). We shall show that any transformation of this form maps circles to circles and straight lines to straight lines. Let z lie on a circle or a straight line. Then z satisfies the equation

$$Az\bar{z} + \bar{B}z + B\bar{z} + C = 0 \qquad (3)$$

(with $A = 0$ if z lies on a straight line, and $A \neq 0$, $|B|^2 - AC > 0$ if z lies on a circle). Let $w = az + b$; then

$$z = \frac{w - b}{a} \quad (a \neq 0),$$

and so substituting in (3), we obtain

$$A\frac{(w - b)(\bar{w} - \bar{b})}{a\bar{a}} + \bar{B}\frac{(w - b)}{a} + B\frac{(\bar{w} - \bar{b})}{\bar{a}} + C = 0,$$

which simplifies to

$$Aw\bar{w} + (\bar{a}\bar{B} - \bar{b}A)w + (aB - bA)\bar{w} + (b\bar{b}A - \bar{a}b\bar{B} - a\bar{b}B + a\bar{a}C) = 0.$$

If $A \neq 0$, then the validity condition (2) holds, for

$$\begin{aligned} (aB - bA)(\bar{a}\bar{B} - \bar{b}A) - A(b\bar{b}A - \bar{a}b\bar{B} - a\bar{b}B + a\bar{a}C) &= |a|^2(|B|^2 - AC) \\ &> 0, \quad \text{since } |B|^2 > AC. \end{aligned}$$

Thus if z lies on a circle, so does w. If z lies on a straight line (that is, $A = 0$), so does w. So the transformation $z \longrightarrow az + b$ maps circles and straight lines to circles and straight lines.

We next consider inversion. If

$$Az\bar{z} + \bar{B}z + B\bar{z} + C = 0,$$

and $w = 1/z$, where $z \neq 0$, then

$$\frac{A}{w\bar{w}} + \frac{\bar{B}}{w} + \frac{B}{\bar{w}} + C = 0,$$

and so

$$Cw\bar{w} + Bw + \bar{B}\bar{w} + A = 0.$$

Thus w lies on a circle if $C \neq 0$, and on a straight line if $C = 0$. So under inversion, circles and straight lines map to circles and straight lines. More explicitly:

a circle not passing through the origin ($A \neq 0, C \neq 0$) maps to a circle not passing through the origin;

a circle passing through the origin ($A \neq 0, C = 0$) maps to a straight line not passing through the origin;

a straight line not passing through the origin ($A = 0, C \neq 0$) maps to a circle passing through the origin;

a straight line passing through the origin ($A = 0, C = 0$) maps to a straight line passing through the origin.

It follows from these considerations that any bilinear transformation maps circles and straight lines to circles and straight lines.

There remain one or two points to be cleared up.

The first is that we have shown that if z lies on a circle, say, and $w = \dfrac{az + b}{cz + d}$ is the image of z under a bilinear transformation, then w lies on a curve C, which for simplicity of exposition we take to be a circle, too. What we have *not* shown is that every point on the circle C is the image of a point on the original circle. In other words, we have not yet shown that the bilinear transformation maps one circle *onto* the other. That the mapping will be onto, in general, follows from the fact that the bilinear transformation has an inverse. The only exceptional case occurs when $c \neq 0$ and the "image" curve passes through the point a/c, since this point is not the image of any point in $\mathbf{C} - \{-d/c\}$. So, for example, in the case of inversion, for which $a/c = 0$, a straight line not passing through the origin is mapped to a circle passing through the origin; but the straight line is not mapped *onto* the circle, because the origin is not the image, under inversion, of any point in $\mathbf{C} - \{0\}$. However, every other point on the circle is the image of a point on the straight line. Inversion maps the straight line onto a circle with one point missing. Incidentally, this kind of problem cannot arise with a bilinear transformation for which $c = 0$, since such a transformation maps $\mathbf{C}$ one-one onto $\mathbf{C}$.

A second point that needs to be cleared up is that we do not know as yet which particular kinds of circle and straight line are mapped to circles, and which to straight lines, under a bilinear transformation $z \longrightarrow \dfrac{az + b}{cz + d}$ for which $c \neq 0$. The best approach to this problem is to observe that a circle is a bounded set, whereas a straight line is not bounded. Now if $c \neq 0$,

$$\frac{az + b}{cz + d} = \frac{a}{c} - \frac{ad - bc}{c} \cdot \frac{1}{cz + d},$$

as we showed earlier (page 23). If C is a curve passing through $-d/c$, then

$$\left\{\frac{1}{cz + d} : z \in C\right\}$$

is unbounded, and so

$$\left\{\frac{az+b}{cz+d}: z \in C\right\}$$

is also unbounded, and the image of C must therefore be a straight line. If C does not pass through $-d/c$, then there is a positive number δ such that $|z + d/c| \geqslant \delta$ for all z in C, since C is a closed set. Thus for z in C,

$$\left|\frac{az+b}{cz+d}\right| \leq \left|\frac{a}{c}\right| + \left|\frac{ad-bc}{c^2}\right| \cdot \frac{1}{\delta},$$

and so the set

$$\left\{\frac{az+b}{cz+d}: z \in C\right\}$$

is bounded. The image of C in this case must be a circle. Applying the same argument to the inverse of the bilinear transformation, we can show that a circle or straight line passing through a/c is the image of a straight line, whereas one not passing through a/c is the image of a circle. We conclude that under a bilinear transformation $f: z \longrightarrow \dfrac{az+b}{cz+d}$ with $c \neq 0$,

> a circle not passing through $-d/c$ is mapped onto a circle not passing through a/c;
>
> a circle passing through $-d/c$ is mapped onto a straight line not passing through a/c;
>
> a straight line not passing through $-d/c$ is mapped onto a circle passing through a/c, but with the point a/c missing;
>
> a straight line passing through $-d/c$ is mapped onto a straight line passing through a/c, but with the point a/c missing.

In Fig. 14 we have illustrated the third of the above cases. Sketch the other three cases for yourself.

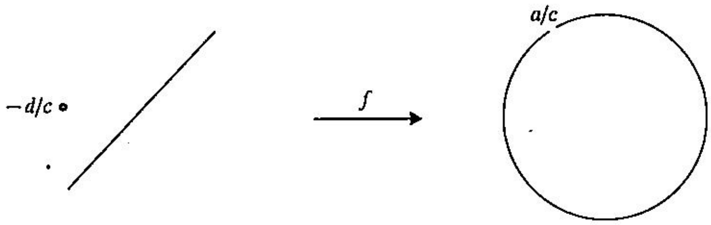


Fig. 14

There are various techniques which may be used to find the image of a particular circle or straight line under a given bilinear transformation. We explain them with the aid of some examples.

Example 1

Find the image of the real axis under the bilinear transformation $z \longrightarrow \dfrac{1+iz}{1-iz}$.

Solution

Since the image is either a circle or straight line, it is enough to know three points on it to determine it completely—in principle, at any rate. In fact, if the image is a straight line, two points will suffice. In this case, however, since the real axis does not pass through the point $-i$ (the singularity of the transformation) the image is a circle (with one point omitted). We can calculate very simply the images of the points $-1, 0, 1$ of the real axis: they are respectively $-i, 1, i$.

The image curve is a circle passing through these three points. We recognize it as the unit circle (with the point -1 omitted).

This method has its limitations, the main one being the difficulty of finding the equation of the circle through the three image points, if they are at all complicated. We have already used, in the simple cases of the basic transformations, a method which leads directly to an equation for the image curve. We now give an example of its use in a more complicated situation.

Example 2

Find the image of the circle $|z - i| = 1$ under the bilinear transformation $z \longrightarrow \dfrac{z+1}{z-1}$.

Solution

We first write the equation of the circle in the form

$$(z - i)(\bar{z} + i) = 1,$$

that is

$$z\bar{z} + iz - i\bar{z} = 0.$$

Now if $w = \dfrac{z+1}{z-1}$, then $z = \dfrac{w+1}{w-1}$. (Check this for yourself, if it surprises you.)

So

$$\frac{(w+1)(\bar{w}+1)}{(w-1)(\bar{w}-1)} + i\left(\frac{w+1}{w-1}\right) - i\left(\frac{\bar{w}+1}{\bar{w}-1}\right) = 0,$$

that is,

$$(w+1)(\bar{w}+1) + i(w+1)(\bar{w}-1) - i(\bar{w}+1)(w-1) = 0,$$

and so,

$$w\bar{w} + (1-2i)w + (1+2i)\bar{w} + 1 = 0.$$

Now since $(1-2i)(1+2i) > 1$, the equation

$$w\bar{w} + (1-2i)w + (1+2i)\bar{w} + 1 = 0$$

is the equation of a circle, as we showed earlier. If we require to know the centre and radius of the circle we may use the formulas given on page 25 or proceed by "completing the square" as follows:

$$\begin{aligned} &w\bar{w} + (1-2i)w + (1+2i)\bar{w} \\ &\quad = \big(w + (1+2i)\big)\big(\bar{w} + (1-2i)\big) - (1+2i)(1-2i) \\ &\quad = |w + (1+2i)|^2 - 5. \end{aligned}$$

Thus $w\bar{w} + (1-2i)w + (1+2i)\bar{w} + 1 = 0$ if and only if

$$|w + (1+2i)|^2 = 4.$$

The image is thus the circle centre $-(1+2i)$, radius 2.

The drawback with this method is the amount of computation involved in all but simple cases. One further, more elegant method, requires a little introduction since it produces the equation of the image curve in an unusual form.

Let a and b be two complex numbers, with $a \neq b$, and consider the image under the bilinear transformation $z \longrightarrow \dfrac{bz-a}{z-1}$ of the circle $|z| = k$ (where k is real and positive, of course). Now if $w = \dfrac{bz-a}{z-1}$, then $z = \dfrac{w-a}{w-b}$, and so if $|z| = k$, then $\left|\dfrac{w-a}{w-b}\right| = k$. It follows that the set $\left\{w : \left|\dfrac{w-a}{w-b}\right| = k\right\}$ defines a circle or a straight line—in fact a circle if $k \neq 1$, a straight line if $k = 1$, when the original circle passes through the singularity of the bilinear transformation, which occurs at the point 1. (See Figs. 15 and 16.)

The form of the image set $\left\{z: \left|\frac{z-a}{z-b}\right| = k\right\}$ is worthy of inspection.

That the equation $|z - a| = |z - b|$ defines a straight line is clear geometrically too—it says, after all, that the distance of z from a is the same as its distance from b, and so it lies on the perpendicular bisector of the line segment $[a, b]$. Our observation that $\left|\frac{z-a}{z-b}\right| = k$ defines a circle when $k \neq 1$ establishes the following interesting geometrical fact: a point which moves so that its distance from one fixed point is a multiple k of its distance from another fixed point describes a circle, if $k \neq 1$. Notice that if a point z satisfies $|z - a| = k|z - b|$ then so does its mirror image z' in the line segment $[a, b]$, since evidently $|z' - a| = |z - a|$ and $|z' - b| = |z - b|$. Thus the line segment $[a, b]$ is an axis of symmetry for the circle, or in other words one of its diameters. This will be useful when it comes to finding the centre and radius of such a circle, as we illustrate in the next example.

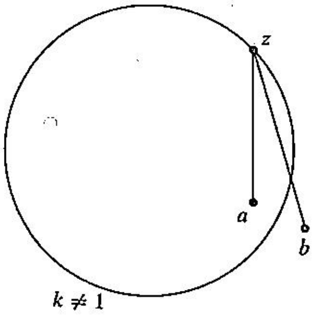


Fig. 15

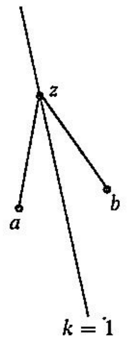


Fig. 16

Example 3

Find the image of the circle $|z - i| = 1$ under the bilinear transformation $z \longrightarrow \frac{z-i}{z+i}$.

Solution

If $w = \frac{z-i}{z+i}$, then $z = i \cdot \frac{1+w}{1-w}$, and so

$$|z - i| = \left|\frac{1+w}{1-w} - 1\right| = \left|\frac{2w}{1-w}\right|.$$

Thus if $|z - i| = 1$, then $\left|\frac{2w}{1-w}\right| = 1$, or equivalently $\left|\frac{w}{w-1}\right| = \frac{1}{2}$. Thus from our discussion above, w lies on a circle, and there is a diameter of the circle along the real axis. We find the ends of that diameter by solving the equation $\left|\frac{w}{w-1}\right| = \frac{1}{2}$ for real w. This gives $\frac{w}{w-1} = \pm\frac{1}{2}$, and so the ends of the diameter are the points -1 and $\frac{1}{3}$. The image circle thus has centre $-\frac{1}{3}$ and radius $\frac{2}{3}$.

We have illustrated three methods of finding the images of circles and straight lines under bilinear transformations. Notice that none of them involves setting $z = x + iy$. To do that leads only to impossible calculations, in general.

Images of Regions

We often need to find the image of a region, usually one bounded by a circle or straight line, under a bilinear transformation. There are basically two methods for finding the image of such a region. The first is algebraic, and is based on the fact that a region bounded by a circle or straight line is defined by an inequality. It is demonstrated in the following example.

Example 4

Show that the bilinear transformation $z \longrightarrow \frac{1+iz}{1-iz}$ maps the upper half-plane $\{z: \operatorname{Im} z > 0\}$ to the unit disc.

Solution

If $w = \dfrac{1+iz}{1-iz}$, where $\operatorname{Im} z > 0$, then $z = i\cdot\dfrac{1-w}{1+w}$, and so

$$\begin{aligned}\operatorname{Im} z &= \operatorname{Im}\left(i\cdot\frac{1-w}{1+w}\right)\\ &= \operatorname{Re}\left(\frac{1-w}{1+w}\right) > 0.\end{aligned}$$

But

$$\begin{aligned}\operatorname{Re}\left(\frac{1-w}{1+w}\right) &= \operatorname{Re}\left(\frac{(1-w)(1+\bar{w})}{(1+w)(1+\bar{w})}\right)\\ &= \operatorname{Re}\left(\frac{1-|w|^2-2i\operatorname{Im} w}{|1+w|^2}\right)\\ &= \frac{1-|w|^2}{|1+w|^2}.\end{aligned}$$

Thus if $\operatorname{Re}\left(\dfrac{1-w}{1+w}\right) > 0$, then $|w| < 1$. We conclude that the upper half-plane is mapped to the unit disc.

Notice that we did not show in Example 4 that the unit disc is the *image* of the upper half-plane under the bilinear transformation, only that it contains the image. It would be surprising if the unit disc were not the image of the upper half-plane since, as we saw in Example 1, the real axis is mapped to the unit circle (with the point -1 missing) under this bilinear transformation. These curves are the boundaries of the regions in question.

These observations suggest another, more topologically based, approach to the problem of Example 4.

Second Solution to Example 4

The bilinear transformation maps $\mathbf{C} - \{-i\}$ one-one onto $\mathbf{C} - \{-1\}$, and it maps the real axis onto the unit circle (with the point -1 missing). Now the upper and lower half-planes are mapped, by the bilinear transformation, to regions whose union is the complement in $\mathbf{C}$ of the unit circle. There are just two regions whose union is the complement in $\mathbf{C}$ of the unit circle: its inside and its outside. Thus the upper half-plane is mapped onto either the inside or the outside of the unit circle. To show that it actually maps onto the inside, we can do one of two things. The first is to observe that since the singularity of the transformation occurs in the lower half-plane, it is that region which is mapped onto the unbounded region outside the unit circle. Alternatively, we may choose just one point in the upper half-plane and calculate its image. If the image of this one point lies inside the unit circle, then the whole region must be mapped onto the inside of the unit circle. The point i is the obvious candidate. It is mapped to 0, which lies inside the unit circle.

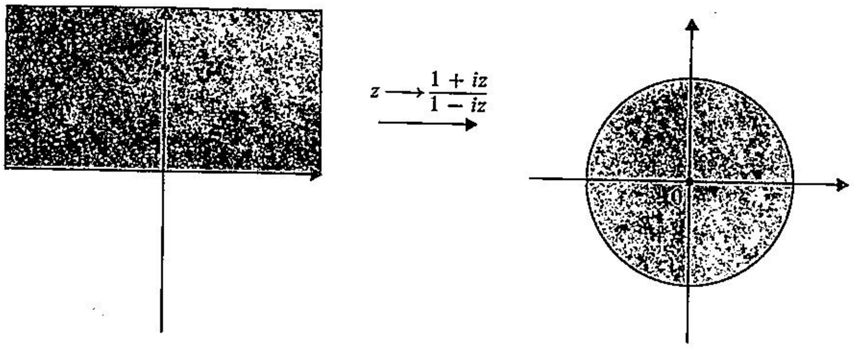


Fig. 17

The rest of this section is devoted to some examples, of a more general character than the previous ones, chosen partly because they illustrate similar points, but also because they establish results useful later in the unit.

Example 5

Show that a bilinear transformation of the form $z \longrightarrow \dfrac{az+b}{\bar{b}z+\bar{a}}$ maps the inside of the unit circle onto the inside or outside of the unit circle according as $|b| < |a|$ or $|b| > |a|$.

Solution

We first show that a bilinear transformation of the given type maps the unit circle to itself. Remembering that if α is any complex number then $|\alpha/\bar{\alpha}| = 1$, we can easily show that the images of the points 1, -1 and i all lie on the unit circle. The image of 1 is

$$\frac{a+b}{\bar{b}+\bar{a}} = \frac{a+b}{\bar{a}+\bar{b}};$$

the image of -1 is

$$\frac{-a+b}{-\bar{b}+\bar{a}} = -\frac{a-b}{\bar{a}-\bar{b}};$$

the image of i is

$$\frac{ai+b}{\bar{b}i+\bar{a}} = i\cdot\frac{a-ib}{\bar{a}-i\bar{b}}.$$

Thus the images of the points 1, -1 and i all have modulus 1, and so lie on the unit circle.

So the image of the inside of the unit circle is either the inside, or the outside, of the unit circle. If $a \neq 0$, the image of 0 is $b/\bar{a}$, which lies inside the unit circle if $|b| < |a|$, and outside if $|b| > |a|$. If $a = 0$, so that $|b| > |a|$ necessarily, the image of the point $\frac{1}{2}$ is $2b/\bar{b}$, which lies outside the unit circle.

Example 6

Find all the bilinear transformations which map the unit circle onto itself and leave the origin fixed, that is the image of the origin is the origin.

If the bilinear transformation $z \longrightarrow \dfrac{az+b}{cz+d}$ leaves 0 invariant, then $b = 0$. If $z \longrightarrow \dfrac{az}{cz+d}$ maps the unit circle to itself, then in particular the images of the points 1 and -1 must lie on the unit circle, and so

$$\left|\frac{a}{c+d}\right| = \left|\frac{a}{-c+d}\right| = 1.$$

Thus $|c+d| = |c-d|$, that is

$$(c+d)(\bar{c}+\bar{d}) = (c-d)(\bar{c}-\bar{d}),$$

and so

$$c\bar{d} + \bar{c}d = 0. \tag{4}$$

The images of the points i and $-i$ must also lie on the unit circle, so

$$\left|\frac{a}{ci+d}\right| = \left|\frac{a}{-ci+d}\right| = 1.$$

Thus $|ci + d| = |ci - d|$, that is

$$(ci + d)(-\bar{c}i + \bar{d}) = (ci - d)(-\bar{c}i - \bar{d}),$$

and so

$$c\bar{d} - \bar{c}d = 0. \tag{5}$$

It follows from (4) and (5) that $c\bar{d} = 0$. It cannot be the case that $\bar{d} = 0$, for then $ad - bc$ would be zero; thus $c = 0$. So the bilinear transformation is $z \longrightarrow az/d$, and we must have $|a/d| = 1$. The transformation is therefore a rotation $z \longrightarrow \beta z$ with $|\beta| = 1$ (as we might well have guessed).

Example 7

Find all the bilinear transformations which map the inside of the unit circle onto itself.

If α lies inside the unit circle, so that $|\alpha| < 1$, then the transformation $z \longrightarrow \dfrac{z - \alpha}{1 - \bar{\alpha}z}$ maps the inside of the unit circle onto itself, by Example 5; also it maps α to 0. So to solve the problem, it is enough to find all bilinear transformations which map the inside of the unit circle onto itself, and leave the origin fixed. For if f maps the inside of the unit circle onto itself, and $f(z) = \alpha$, and if $g(z) = \dfrac{z - \alpha}{1 - \bar{\alpha}z}$, then $g \circ f$ maps the inside of the unit circle onto itself, and leaves the origin fixed. We can recover the formula for f if we can find $g \circ f$. We have found the form of $g \circ f$ in Example 6. Thus the general form of a bilinear transformation which maps the inside of the unit circle onto itself is obtained by solving

$$\frac{f(z) - \alpha}{1 - \bar{\alpha}f(z)} = \beta z$$

for $f(z)$. Let $\beta = e^{i\theta}$; then

$$f(z) = \frac{e^{i\theta}z + \alpha}{\bar{\alpha}e^{i\theta}z + 1}.$$

This transformation is of the form $z \longrightarrow \dfrac{az + b}{\bar{b}z + \bar{a}}$, where $a = e^{i\theta/2}$ and $b = \alpha e^{-i\theta/2}$. Note that $|b| = |\alpha| < 1 = |a|$.

So the bilinear transformations which map the inside of the unit circle onto itself are those of the form $z \longrightarrow \dfrac{az + b}{\bar{b}z + \bar{a}}$, where $|b| < |a|$.

Example 8

Let C_1 and C_2 be two circles which are not concentric and do not intersect, and such that C_1 lies inside C_2. Show that there is a bilinear transformation which maps the region lying between C_1 and C_2 to an annulus.

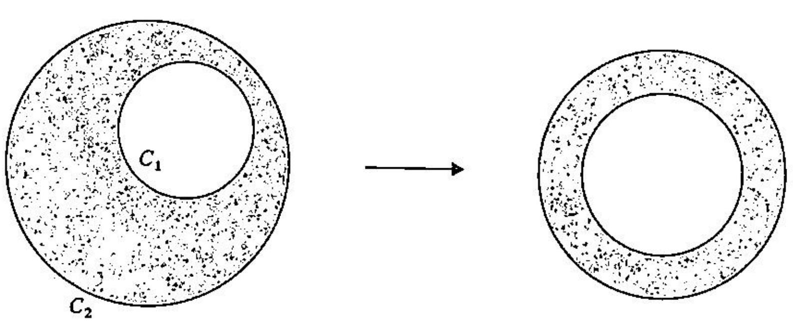


Fig. 18

Solution

When faced with a rather general problem of this kind, it often helps to begin by using a bilinear transformation to move the figure into a standard position. In this case we can move the centre of C_2 to the origin, by carrying out a translation, and then adjust the radius of C_2 so that it becomes 1, by carrying out a magnification; we then arrange that the centre of C_1 lies on the positive real axis, by carrying out a rotation about the origin. The composition of these transformations is a bilinear transformation; if we can find a bilinear transformation which has the desired effect on the region between the circles in the standardized position, then the composition of the two bilinear transformations solves the general problem. So we may as well assume that C_2 is the unit circle; and since translations, magnifications and rotations do not interchange inside and outside, that C_1 lies inside C_2 and has its centre at the point x on the real axis, where $0 < x < 1$. If ρ is the radius of C_1 we must have $x + \rho < 1$.

The idea is to try to find a bilinear transformation which leaves C_2 where it is, as the unit circle, but turns C_1 into a circle inside C_2 with centre at 0. We know all the bilinear transformations which map the unit circle onto itself, and its inside onto its inside: by Examples 6 and 7, they are the transformations of the form $z \longrightarrow \beta\left(\frac{z-\alpha}{1-\bar{\alpha}z}\right)$ where $|\beta| = 1$ and $0 \leqslant |\alpha| < 1$. We shall show, in fact, that we can find a bilinear transformation with $\beta = 1$ and α real and positive which maps the circle C_1 to one with its centre at 0. It isn't very surprising that we can impose these restrictions on the bilinear transformation. The standardization of the problem has made most of the variables real so we might expect α to be real; the final position is unchanged by rotation about the origin, so we might expect β to be 1.

The circle C_1 is given by $|z - x| = \rho$. The transformation in question is $z \longrightarrow \frac{z-\alpha}{1-\alpha z}$. Now if $w = \frac{z-\alpha}{1-\alpha z}$, then $z = \frac{w+\alpha}{\alpha w+1}$, and so C_1 is transformed to

$$\left|\frac{w+\alpha}{\alpha w+1} - x\right| = \rho,$$

that is

$$|(1-\alpha x)w + (\alpha - x)| = \rho|\alpha w + 1|,$$

and so

$$\left|w + \frac{\alpha - x}{1-\alpha x}\right| = \frac{\rho\alpha}{1-\alpha x}\left|w + \frac{1}{\alpha}\right|$$

since $0 < x < 1$ and $0 \leqslant \alpha < 1$, $1 - \alpha x > 0$.

This is the equation of a circle in the form given in Example 3. The real axis is a diameter. We must show that we can choose α so that the circle has its centre at 0. The points at which it cuts the real axis, which are the ends of a diameter, are the solutions of

$$w + \frac{\alpha - x}{1-\alpha x} = \pm\frac{\rho\alpha}{1-\alpha x}\left(w + \frac{1}{\alpha}\right),$$

which are $\frac{(x+\rho)-\alpha}{1-\alpha(x+\rho)}$ and $\frac{(x-\rho)+\alpha}{1-\alpha(x-\rho)}$. So the circle has centre 0 provided

$$\frac{(x+\rho)-\alpha}{1-\alpha(x+\rho)} + \frac{(x-\rho)+\alpha}{1-\alpha(x-\rho)} = 0.$$

This will be the case if

$$[(x+\rho)-\alpha][1-\alpha(x-\rho)] + [(x-\rho)+\alpha][1-\alpha(x+\rho)] = 0$$

or equivalently

$$x - (1 + x^2 - \rho^2)\alpha + x\alpha^2 = 0.$$

This is a quadratic equation for α.

Now $x + \rho < 1$, $\rho > 0$ and $0 < x < 1$, so

$$0 < \rho < 1 - x.$$

Thus $\rho^2 < (1 - x)^2$, and so $0 < 2x < x^2 + 1 - \rho^2$; on squaring, we obtain

$$4x^2 < (x^2 + 1 - \rho^2)^2.$$

Thus the quadratic equation has two distinct real roots. Their product is 1, as is clear from the equation; it follows that one of them is less than 1 in absolute value. Consider now the bilinear transformation $z \longrightarrow \dfrac{z - \alpha_1}{1 - \alpha_1 z}$ where α_1 is that (real) root of the quadratic equation $x\alpha^2 - (x^2 + 1 - \rho^2)\alpha + x = 0$ whose absolute value is less than 1. The bilinear transformation maps C_2 (the unit circle) onto itself, and the inside of C_2 onto itself; it maps C_1 onto a circle concentric with C_2, and the inside of C_1 onto the inside of the image circle. Thus the region between C_1 and C_2 is mapped onto the annulus lying between the image of C_1 and C_2.

Self-Assessment Question 3

Answer the following questions about the bilinear transformation $f(z) = \dfrac{1 + iz}{1 - iz}$.

(i) Calculate $f(i)$ and $f(0)$.

(ii) What is the domain of f? In other words, where does its singularity lie?

(iii) Is the image under f of the imaginary axis a circle or a straight line?

(iv) What is the image under f of the imaginary axis? (You have already calculated the images of two points on it.)

(v) What is the image under f of the right half-plane $\{z : \operatorname{Re} z > 0\}$?

Solution

(i) $f(i) = 0, \quad f(0) = 1.$

(ii) f has its singularity at $-i$, so its domain is $\mathbf{C} - \{-i\}$.

(iii) Since the imaginary axis passes through $-i$, the singularity of f, its image is a straight line with one point missing.

(iv) The image of the imaginary axis is a straight line passing through $f(i)$ and $f(0)$, that is, through 0 and 1. It is thus the real axis with the point -1 missing.

(v) The image of the right half-plane is either the upper half-plane or the lower half-plane. To find out which, it is enough to calculate the image of one point in the right half-plane. The point 1 is a suitable point; its image is $\dfrac{1 + i}{1 - i} = i$. Thus the image under f of the right half-plane is the upper half-plane.

Summary

We have described the main geometric properties of bilinear transformations. These are:

(i) a bilinear transformation is an analytic, conformal transformation of the plane, or plane with a point removed;

(ii) the family of bilinear transformations contains all the important transformations of geometry, namely isometries, similarities and inversion (but not reflection);

(iii) a bilinear transformation maps circles and straight lines to circles and straight lines.

13.4 PROBLEMS

1. Prove that

   (i) a bilinear transformation is conformal;

   (ii) the inverse of the bilinear transformation $z \longrightarrow \dfrac{az+b}{cz+d}$ is the transformation $z \longrightarrow \dfrac{dz-b}{-cz+a}$;

   (iii) if $c \neq 0$, the image of the region $\mathbf{C} - \{-d/c\}$ under the transformation $z \longrightarrow \dfrac{az+b}{cz+d}$ is the region $\mathbf{C} - \{a/c\}$;

   (iv) if $ad - bc = 0$, and at least one of c and d is non-zero, then the function $z \longrightarrow \dfrac{az+b}{cz+d}$ is a constant function;

   (v) the composition of two bilinear transformations is a bilinear transformation;

   (vi) the coefficients a, b, c, d of a bilinear transformation $z \longrightarrow \dfrac{az+b}{cz+d}$ can be multiplied by a non-zero complex number without affecting the transformation, and thus we may assume that $ad - bc = 1$.

2. Let r be a real number, $0 < r < 1$. Let $\theta_r(z) = \dfrac{z-r}{rz-1}$. Show that θ_r maps the unit disc $\{z:|z| < 1\}$ onto itself. Find $\theta_r(r)$ and $\theta_r(0)$. Find the inverse of θ_r.

3. Find a bilinear transformation which maps the circle $|z| = r_1$ onto the circle $|z| = r_2$; then find one which maps the circle $|z - z_0| = r$ onto the circle $|z| = r$. Use your answers to construct a bilinear transformation which maps the circle $|z - z_1| = r_1$ onto the circle $|z - z_2| = r_2$. Does your transformation map the inside of the first circle onto the inside or the outside of the second? Find a transformation which does the opposite.

4. Show that if a, b, c, d are all real, then the bilinear transformation $z \longrightarrow \dfrac{az+b}{cz+d}$ maps the upper half-plane $\{z:\operatorname{Im} z > 0\}$ onto itself, or onto the lower half-plane $\{z:\operatorname{Im} z < 0\}$, according as $ad - bc$ is positive or negative.

   Find all bilinear transformations which map the upper half-plane onto itself.

5. The expression $\dfrac{az+b}{cz+d}$ contains four arbitrary constants a, b, c and d, but we need only three ratios to specify uniquely the corresponding bilinear transformation. For, since $ab - cd \neq 0$, at least one of c and d is non-zero, and supposing that $c \neq 0$, the transformation is

$$T:z \longrightarrow \frac{\dfrac{a}{c}z + \dfrac{b}{c}}{z + \dfrac{d}{c}}.$$

   Since we require only three ratios to determine T we might suspect that three points and their images are sufficient to specify T. We shall show in this problem that this is indeed the case.

Suppose that

$$w_k = T(z_k) = \frac{az_k + b}{cz_k + d}, \quad k = 1, 2, 3;$$

then our first thought might be to solve these three equations for the three ratios which determine T. However, it is probably easier to proceed as follows.

We shall see later in this unit that it is sensible to talk about "the point at infinity", but for the moment let us use the notion in an intuitive sense to motivate the following approach.

The transformation $S: z \longrightarrow \frac{z - z_1}{z - z_2}$ maps the points z_1 and z_2 to zero and "infinity" respectively; the transformation $S_1: w \longrightarrow \frac{w - w_1}{w - w_2}$ maps the points w_1 and w_2 also to zero and "infinity". Thus we might expect that $S_1^{-1} \circ S$ maps z_1 to w_1 and z_2 to w_2.

But if $w = S_1^{-1}(S(z))$ then $S_1(w) = S(z)$, so that

$$\frac{w - w_1}{w - w_2} = \frac{z - z_1}{z - z_2},$$

and we are therefore led to examine transformations from z to w defined by an equation of the form

$$\frac{\alpha w + \beta}{\gamma w + \delta} = \frac{\alpha_1 z + \beta_1}{\gamma_1 z + \delta_1}.$$

(i) Consider the transformation $F: z \longrightarrow w$ defined by

$$\frac{\alpha w + \beta}{\gamma w + \delta} = \frac{\alpha_1 z + \beta_1}{\gamma_1 z + \delta_1},$$

$$\text{where } \alpha\delta - \beta\gamma \neq 0, \alpha_1\delta_1 - \beta_1\gamma_1 \neq 0. \qquad (1)$$

Show that F is a bilinear transformation.

(ii) By using the fact that

$$w_k = T(z_k), \quad k = 1, 2, 3,$$

show that the equation of the form (1) which specifies T uniquely is

$$\frac{(w - w_1)(w_3 - w_2)}{(w - w_2)(w_3 - w_1)} = \frac{(z - z_1)(z_3 - z_2)}{(z - z_2)(z_3 - z_1)}.$$

6. Let z_1, z_2, z_3, z_4 be four distinct complex numbers. We call

$$\frac{(z_1 - z_2)(z_3 - z_4)}{(z_2 - z_3)(z_4 - z_1)}$$

the *cross-ratio* of z_1, z_2, z_3, z_4. (Note that the order is important: z_2, z_1, z_3, z_4 have a different cross-ratio from z_1, z_2, z_3, z_4.) Show that if f is a bilinear transformation the cross-ratio of $f(z_1), f(z_2), f(z_3), f(z_4)$ is the same as the cross-ratio of z_1, z_2, z_3, z_4. (Work directly; do not try to use Problem 5.)

Show that if four points z_1, z_2, z_3, z_4 lie on a line, then their cross-ratio is a real number. Show that the same thing is true of four points which lie on a circle. (Use Problem 5 here.)

7. A point α is called a *fixed point* of a bilinear transformation f if $f(\alpha) = \alpha$.

   Show that in general a bilinear transformation has two distinct fixed points. Which bilinear transformations have no fixed points? Show that the bilinear transformation $z \longrightarrow \dfrac{az+b}{cz+d}$ has just one fixed point if and only if either

   $$(d-a)^2 = -4bc \quad \text{and} \quad c \neq 0;$$

   or

   $$c = 0 \quad \text{and} \quad a \neq d.$$

   Show that there is one and only one bilinear transformation which has more than two fixed points.

The following two problems are for interest only; do not attempt them unless you are interested and have the time.

8. Here is an example of a geometric use of bilinear transformations.

   Let C_1 and C_2 be two non-intersecting circles, with C_1 inside C_2, as shown in Fig. 19. Further circles $S_1, S_2, \ldots$ are drawn as shown. S_1 touches C_1 and C_2; S_2 touches C_1, C_2 and S_1; S_3 touches C_1, C_2 and S_2; and so on. We say that the chain of circles $S_1, S_2, \ldots$ *closes* if there is a circle S_n which touches C_1, C_2, S_{n-1} and S_1. (The chain in Fig. 19 closes.) Prove that if the chain of circles closes for one position of the circle S_1 then it closes for all positions of S_1. (Hint: Example 8, Section 13.3).

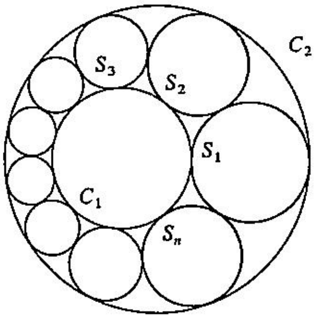


Fig. 19

9. Show that an isometry is necessarily of one or the other of the forms

   $$z \longrightarrow az + b, \quad \text{where } |a| = 1,$$

   or

   $$z \longrightarrow a\bar{z} + b, \quad \text{where } |a| = 1.$$

Solutions

1. (i) Let $f(z) = \dfrac{az+b}{cz+d}$ be a bilinear transformation. Then

   $$f'(z) = \frac{a(cz+d) - c(az+b)}{(cz+d)^2} = \frac{ad-bc}{(cz+d)^2}.$$

   So $f'(z) \neq 0$ (since $ad - bc \neq 0$), and f is conformal.

   (ii) The simplest thing to do is to compose the two transformations and see if we obtain the identity function.

   $$\frac{a\left(\dfrac{dz-b}{-cz+a}\right)+b}{c\left(\dfrac{dz-b}{-cz+a}\right)+d} = \frac{a(dz-b)+b(-cz+a)}{c(dz-b)+d(-cz+a)}$$

   $$= \frac{(ad-bc)z}{-bc+ad}$$

   $$= z.$$

   Alternatively, we may solve the equation $w = \dfrac{az+b}{cz+d}$ for z. This gives

   $$w(cz+d) = az+b,$$

   that is,

   $$z(cw-a) = -dw+b,$$

   and so

   $$z = \frac{-dw+b}{cw-a}.$$

   Oh dear, the signs seem to be wrong. (But see part (vi).)

(iii) Given any complex number w, except a/c, there is a complex number z such that $\frac{az+b}{cz+d} = w$, namely $z = \frac{dw-b}{-cw+a}$.

(iv) If $ad - bc = 0$ and $c \neq 0$, then $b = ad/c$, and

$$\frac{az+b}{cz+d} = \frac{acz+ad}{c(cz+d)} = \frac{a}{c}.$$

If $ad - bc = 0$ and $d \neq 0$, then $a = bc/d$, and

$$\frac{az+b}{cz+d} = \frac{bcz+bd}{d(cz+d)} = \frac{b}{d}.$$

(v) Let $f_1(z) = \frac{a_1z+b_1}{c_1z+d_1}$ and $f_2(z) = \frac{a_2z+b_2}{c_2z+d_2}$ be two bilinear transformations; then

$$\begin{aligned}(f_1 \circ f_2)(z) &= \frac{a_1\left(\frac{a_2z+b_2}{c_2z+d_2}\right)+b_1}{c_1\left(\frac{a_2z+b_2}{c_2z+d_2}\right)+d_1}\\ &= \frac{a_1(a_2z+b_2)+b_1(c_2z+d_2)}{c_1(a_2z+b_2)+d_1(c_2z+d_2)}\\ &= \frac{(a_1a_2+b_1c_2)z+(a_1b_2+b_1d_2)}{(c_1a_2+d_1c_2)z+(c_1b_2+d_1d_2)}.\end{aligned}$$

This is a bilinear transformation because

$$\begin{aligned}&(a_1a_2+b_1c_2)(c_1b_2+d_1d_2)-(c_1a_2+d_1c_2)(a_1b_2+b_1d_2)\\ &\quad= (a_1d_1-b_1c_1)(a_2d_2-b_2c_2)\\ &\quad\neq 0.\end{aligned}$$

(vi) If λ is a non-zero complex number then

$$\frac{\lambda az+\lambda b}{\lambda cz+\lambda d} = \frac{az+b}{cz+d} \quad \text{for every } z\left(\neq -\frac{d}{c}\right).$$

Thus a, b, c, d and $\lambda a, \lambda b, \lambda c, \lambda d$ define the same bilinear transformation (so $z \longrightarrow \frac{-dz+b}{cz-a}$ is the same transformation as $z \longrightarrow \frac{dz-b}{-cz+a}$, and these two expressions are equally valid ways of writing the inverse of $z \longrightarrow \frac{az+b}{cz+d}$.) Suppose that $ad - bc \neq 1$. Let $1/\lambda$ be a square root of $ad - bc$. Let $a' = \lambda a$, $b' = \lambda b$, and so on. Then a', b', c', d' define the same bilinear transformation as a, b, c, d and

$$a'd' - b'c' = \frac{ad-bc}{\lambda^2} = 1.$$

2. The function θ_r is of the form $z \longrightarrow \frac{az+b}{\bar{b}z+\bar{a}}$ with $a = i$ and $b = -ir$; so by Example 5 of Section 13.3 it maps the unit disc onto itself (since $|b| = r < 1 = |a|$). Also $\theta_r(r) = 0$, $\theta_r(0) = r$, and $\theta_r^{-1}(z) = \frac{-z+r}{-rz+1} = \frac{z-r}{rz-1}$. Thus $\theta_r^{-1} = \theta_r$.

3. The magnification $z \longrightarrow \frac{r_2}{r_1}z$ maps the circle $|z| = r_1$ onto the circle $|z| = r_2$. The translation $z \longrightarrow z - z_0$ maps the circle $|z - z_0| = r$ onto the circle $|z| = r$.

To find a bilinear transformation which maps the circle $|z - z_1| = r_1$ onto the circle $|z - z_2| = r_2$, we proceed as follows. Translate the first circle so that its centre is moved to 0; magnify, to get the radius right; translate the centre to z_2. The bilinear transformation which will do this is the composition $h \circ g \circ f$ where

$$f: z \longrightarrow z - z_1,$$

$$g: z \longrightarrow \frac{r_2}{r_1}z,$$

$$h: z \longrightarrow z + z_2,$$

which is $z \longrightarrow \frac{r_2z + r_1z_2 - r_2z_1}{r_1}$.

This transformation maps z_1 to z_2, so it maps the inside of the first circle to the inside of the second. To obtain a bilinear transformation which interchanges inside and outside, we may insert an inversion at an appropriate stage. For example, the composition $h \circ k \circ g \circ f$, where

$$f: z \longrightarrow z - z_1,$$

$$k: z \longrightarrow \frac{r_1^2}{z},$$

$$g: z \longrightarrow \frac{r_2}{r_1} z,$$

$$h: z \longrightarrow z + z_2,$$

which is $z \longrightarrow \dfrac{z_2 z + r_1 r_2 - z_1 z_2}{z - z_1}$, will have this effect.

4. Since a, b, c, d are real, the image of any point on the real axis under $z \longrightarrow \dfrac{az + b}{cz + d}$ evidently lies on the real axis. The transformation therefore maps the real axis onto itself, and consequently maps the upper half-plane onto either the upper or the lower half-plane. To find out which, we consider the image of a point in the upper half-plane, the point i for example. Now

$$\frac{ai + b}{ci + d} = \frac{(ai + b)(-ci + d)}{c^2 + d^2} = \frac{ac + bd + (ad - bc)i}{c^2 + d^2}.$$

This point lies in the upper half-plane if $ad - bc > 0$, and in the lower half-plane if $ad - bc < 0$.

Suppose that $z \longrightarrow \dfrac{az + b}{cz + d}$ maps the upper half-plane onto itself. Then it must map the real axis onto itself. Suppose first that $c \neq 0$. Since the image of the real axis is a straight line, the real axis itself must pass through the point $-d/c$. Thus $-d/c$ is real. Since the real axis is itself a straight line, its image passes through the point a/c. Thus a/c is real. Finally, the image of the point 0 must lie on the real axis; thus b/d is real. Let $-d/c = k$, $a/c = l$, $b/d = m$, where k, l, m are real. Then

$$d = -kc; a = lc; b = md = -mkc$$

and so

$$\frac{az + b}{cz + d} = \frac{lcz - mkc}{cz - kc} = \frac{lz - mk}{z - k}$$

and the bilinear transformation has real coefficients (or rather may be made to have real coefficients after suitable cancellations). Suppose on the other hand that $c = 0$. We may as well assume that $d = 1$, so the bilinear transformation is $z \longrightarrow az + b$. Then the images of 0 and 1 must be real, so b and $a + b$ must be real, and so a and b must both be real.

We conclude that the bilinear transformations that map the upper half-plane onto itself are precisely those that may be brought to the form $z \longrightarrow \dfrac{az + b}{cz + d}$ with a, b, c, d real and $ad - bc > 0$.

5. (i) Let $S_1: w \longrightarrow \dfrac{\alpha w + \beta}{\gamma w + \delta}$ and $S: z \longrightarrow \dfrac{\alpha_1 z + \beta_1}{\gamma_1 z + \delta_1}$. Then $F = S_1^{-1} \circ S$, and since S and S_1 are bilinear transformations so is F.

(ii) First notice that if z takes a value which makes the numerator on the right of

$$\frac{\alpha w + \beta}{\gamma w + \delta} = \frac{\alpha_1 z + \beta_1}{\gamma_1 z + \delta_1}$$

take the value zero, then w must take the value which makes the numerator on the left zero. Similar remarks apply to the denominator. Thus we can write the equation specifying the transformation T in the form

$$\frac{w - w_1}{w - w_2} = \lambda \frac{z - z_1}{z - z_2}, \tag{2}$$

where λ is a constant which has to be determined. We now use the remaining information, namely $T: z_3 \longrightarrow w_3$; so that

$$\frac{w_3 - w_1}{w_3 - w_2} = \lambda \frac{z_3 - z_1}{z_3 - z_2}. \tag{3}$$

If we divide equation (2) by equation (3), we obtain

$$\frac{(w - w_1)(w_3 - w_2)}{(w - w_2)(w_3 - w_1)} = \frac{(z - z_1)(z_3 - z_2)}{(z - z_2)(z_3 - z_1)},$$

which specifies the bilinear transformation T uniquely.

6. For any pair of points z_1 and z_2, we have

$$\begin{aligned} f(z_1) - f(z_2) &= \frac{az_1 + b}{cz_1 + d} - \frac{az_2 + b}{cz_2 + d} \\ &= \frac{(ad - bc)(z_1 - z_2)}{(cz_1 + d)(cz_2 + d)}, \end{aligned}$$

where $f(z) = \dfrac{az + b}{cz + d}$. Thus

$$\begin{aligned} \frac{\big(f(z_1) - f(z_2)\big)\big(f(z_3) - f(z_4)\big)}{\big(f(z_2) - f(z_3)\big)\big(f(z_4) - f(z_1)\big)} &= \frac{\dfrac{(ad - bc)^2(z_1 - z_2)(z_3 - z_4)}{(cz_1 + d)(cz_2 + d)(cz_3 + d)(cz_4 + d)}}{\dfrac{(ad - bc)^2(z_2 - z_3)(z_4 - z_1)}{(cz_2 + d)(cz_3 + d)(cz_4 + d)(cz_1 + d)}} \\ &= \frac{(z_1 - z_2)(z_3 - z_4)}{(z_2 - z_3)(z_4 - z_1)}. \end{aligned}$$

If z_1, z_2, z_3, z_4 lie on a line, we may write $z_i = a + bt_i$ for $i = 1, 2, 3, 4$, where the t_i are real numbers. Then

$$\begin{aligned} \frac{(z_1 - z_2)(z_3 - z_4)}{(z_2 - z_3)(z_4 - z_1)} &= \frac{b^2(t_1 - t_2)(t_3 - t_4)}{b^2(t_2 - t_3)(t_4 - t_1)} \\ &= \frac{(t_1 - t_2)(t_3 - t_4)}{(t_2 - t_3)(t_4 - t_1)}, \end{aligned}$$

which is real.

Suppose that z_1, z_2, z_3, z_4 lie on a circle. Let w_1, w_2, w_3 be any three distinct collinear points. There is a bilinear transformation which maps z_1, z_2, z_3 to w_1, w_2, w_3 respectively. The image of the circle through z_1, z_2, z_3 is either a circle or a straight line, and passes through w_1, w_2, w_3; it is therefore the straight line on which these points lie. The image of z_4, which we may as well call w_4, lies on the line through w_1, w_2, w_3. The cross-ratio of z_1, z_2, z_3, z_4 is the same as the cross-ratio of w_1, w_2, w_3, w_4; it is therefore real, since those points lie on a line.

7. If $f(z) = \dfrac{az + b}{cz + d}$ and $f(\alpha) = \alpha$, then $\alpha = \dfrac{a\alpha + b}{c\alpha + d}$, and so

$$c\alpha^2 + (d - a)\alpha - b = 0.$$

Conversely, if α is a solution of the quadratic equation

$$cz^2 + (d - a)z - b = 0,$$

it is a fixed point of f. Since in general a quadratic equation has two distinct solutions, there will be in general two fixed points. The special cases occur where there are extra conditions on a, b, c, d. These cases are:

$c \neq 0, (d - a)^2 = -4bc$: one fixed point
$c = 0, d \neq a$: one fixed point
$c = 0, d = a, b \neq 0$: no fixed points
$c = 0, d = a, b = 0$: every point a fixed point.

Thus the only bilinear transformations with no fixed point are those for which $c = 0$, $d = a$, $b \neq 0$, which are the translations. The bilinear transformations with one and only one fixed point are those for which $c \neq 0$ and $(d - a)^2 = -4bc$, and those for which $c = 0$ and $d \neq a$. The only bilinear transformation with more than two fixed points has $b = c = 0$ and $a = d$, and is the identity function.

8.

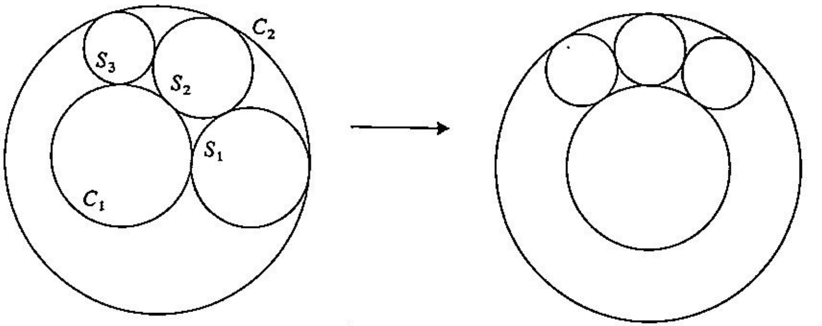


Fig. 20

There is a bilinear transformation which maps C_1 and C_2 to a pair of concentric circles. It maps $S_1, S_2, \ldots$ to circles which touch C_1, C_2, and each one except the first also touches the previous one. In other words, the images of $S_1, S_2, \ldots$ form a chain of circles for the concentric pair of circles which are the images of C_1 and C_2. The original chain closes if and only if the image chain does. Whether or not the chain closes for a pair of concentric circles depends only on the ratio of their radii, not on the position of the initial circle in the chain. Thus if the outer and inner circles are concentric the chain closes for every position of the initial circle if it closes for one position of the initial circle. Thus the same is true for any pair of circles C_1 and C_2.

9. Note first of all that both the given transformations are isometries, and that the composition of isometries is an isometry.

Suppose that f is an isometry. Let $f(0) = b$; then $z \longrightarrow z - b$ is an isometry, and if $f_1(z) = f(z) - b$, then f_1 is an isometry such that $f_1(0) = 0$.

Suppose that $f_1(1) = a$. Then since f_1 is an isometry,

$$\begin{aligned} |f_1(1) - f_1(0)| &= |a| \\ &= |1 - 0| = 1. \end{aligned}$$

Thus $|a| = 1$, and so $z \longrightarrow \dfrac{1}{a}z$ is an isometry. Let $f_2(z) = \dfrac{1}{a}f_1(z)$; then f_2 is an isometry such that $f_2(1) = 1$.

Now we consider $f_2(i)$. Suppose $f_2(i) = u + iv$. Then since $|f_2(i) - f_2(0)| = |i - 0| = 1$, we have

$$u^2 + v^2 = 1;$$

also, since $|f_2(i) - f_2(1)| = |i - 1| = \sqrt{2}$, we have

$$(u - 1)^2 + v^2 = 2.$$

Thus $u^2 - 2u + 1 + v^2 = 2$, and so

$$u = 0, \quad \text{and} \quad v = \pm 1.$$

So, either $f_2(i) = i$, or $f_2(i) = -i$.

Let $f_3(z) = f_2(z)$ if $f_2(i) = i$, and let $f_3(z) = \overline{f_2(z)}$ if $f_2(i) = -i$. Then f_3 is an isometry such that $f_3(0) = 0$, $f_3(1) = 1$ and $f_3(i) = i$. We shall show that f_3 must be the identity function. Take any point $z = x + iy$; let $f_3(z) = w = u + iv$. Then

$$|w| = |f_3(z) - f_3(0)| = |z - 0|;$$

$$|w - 1| = |f_3(z) - f_3(1)| = |z - 1|;$$

and

$$|w - i| = |f_3(z) - f_3(i)| = |z - i|.$$

So

$$u^2 + v^2 = x^2 + y^2;$$

$$(u - 1)^2 + v^2 = (x - 1)^2 + y^2;$$

and

$$u^2 + (v - 1)^2 = x^2 + (y - 1)^2.$$

It follows from the first two of these three equations that

$$-2u + 1 = -2x + 1,$$

and so $u = x$; and from the first and third of the three equations, that

$$-2v + 1 = -2y + 1,$$

and so $v = y$.

Thus $f_3(z) = z$.

So, either $f_2(z) = z$ or $f_2(z) = \bar{z}$, and either $f_1(z) = az$ or $f_1(z) = a\bar{z}$, where $|a| = 1$, and either $f(z) = az + b$ or $f(z) = a\bar{z} + b$, where $|a| = 1$.

13.5 THE POINT AT INFINITY

We have proved the following facts about a bilinear transformation $z \longrightarrow \dfrac{az+b}{cz+d}$, where $c \neq 0$: it maps $\mathbf{C} - \{-d/c\}$ one-one onto $\mathbf{C} - \{a/c\}$; it maps circles to circles in general, but the image of a circle through the point $-d/c$ is a straight line; it maps a straight line to a circle (or straight line, if the original straight line passes through $-d/c$) with a point missing, the point a/c. You were probably left with the feeling that these results are untidy and unsatisfactory, because of the many special cases we have to consider. How much neater it would be if we could simply say that a bilinear transformation maps circles to circles, and be done with it. After all, a straight line is sometimes thought of as a circle of a special kind, one which passes through some mysterious point "at infinity".

In this section we propose to examine this notion of a point at infinity, and show that we can make good mathematical sense of it, and even suggest a way in which "infinity" can be visualized.

* * * * * * * *

We have already used the symbol ∞, but so far only in the context of a limit, and as shorthand for some complicated expression involving ε and N. You may remember that we adopted the following definition in *Unit 5, Cauchy's Theorem I* (page 90):

> we say that f has limit l near ∞, and write $\lim\limits_{z\to\infty} f(z) = l$, if $\lim\limits_{z\to 0} f(1/z) = l$.

In *Unit 8, Singularities,* we had this definition (page 21):

> $\lim\limits_{z\to a} f(z) = \infty$ means that given any positive number N there is a positive number δ such that $|f(z)| > N$ whenever $0 < |z - a| < \delta$.

Now $\lim\limits_{z\to -d/c}\left(\dfrac{az+b}{cz+d}\right) = \infty$, and $\lim\limits_{z\to\infty}\left(\dfrac{az+b}{cz+d}\right) = \dfrac{a}{c}$ (when $c \neq 0$, of course). This suggests the following story, which hangs together very well. We should imagine that there is a point ∞, in addition to all the ordinary points of the complex plane, and that our bilinear transformation maps the point $-d/c$ to the point ∞, and the point ∞ to the point a/c. We can conveniently group together into one set circles and straight lines, distinguishing the straight lines as those curves in the set which pass through ∞. Then our bilinear transformation maps circles to circles. Since $-d/c$ is mapped to ∞, any circle through that point is mapped to a circle through ∞, that is, a straight line. Moreover, a straight line, that is to say a circle through ∞, is mapped to a circle through the image of ∞, namely a/c.

We shall now take the notion of a point at infinity seriously. It will not be enough, clearly, for us just to add a point to the complex plane—that is, simply take the union of the plane and one extra point. The point at infinity must be attached to the plane rather more firmly than that. We shall say below how we define a neighbourhood of the point at infinity. It is the way we do this that specifies how the extra point is to be attached to the plane. Notice that our definition of $\lim\limits_{z\to\infty} f(z)$ above is really based on the following idea of the point at infinity: it is "the image of 0 under the function $z \longrightarrow 1/z$". In making all our definitions we shall be guided by this idea. (To define $\lim\limits_{z\to a} f(z) = \infty$ in this spirit, we have to make the transformation $z \longrightarrow 1/z$ in the range of f. A definition equivalent to the one we gave above is that $\lim\limits_{z\to a} f(z) = \infty$ if $\lim\limits_{z\to a} 1/f(z) = 0$.)

Definition

The **extended complex plane**, $\hat{C}$, is the union of the ordinary complex plane C and one extra point, **the point at infinity**, which is denoted by ∞. In symbols, $\hat{C} = C \cup \{\infty\}$. A set D in $\hat{C}$ is a **neighbourhood of the point at infinity**, or neighbourhood of ∞, if $\{1/z : z \in D, z \neq \infty\}$ is a punctured neighbourhood of 0.

We shall use the symbol z to denote an unspecified point of the extended complex plane. Thus expressions like $z = \infty$ or $z \neq \infty$ will be permitted.

A neighbourhood of ∞ is the outside of a circle centre 0 (Fig. 21). With this notion of a neighbourhood of ∞ at our disposal, we can extend some of the topological ideas of *Unit 2* to subsets of the extended complex plane. For example, a set S in $\hat{C}$ is **open** if for every point in S, a neighbourhood of the point is also contained in S. An open set in $\hat{C}$ which does not contain the point at infinity is merely an open set in C. But an open set in $\hat{C}$ which does contain the point at infinity must also contain a set of the form $\{z : |z| > r\}$. The complement of S in $\hat{C}$ is therefore contained in $\{z : |z| \leqslant r\}$; in other words, it is bounded. The complement of an open set (even one in $\hat{C}$) is **closed**. So the complement of S is closed and bounded, that is, compact. Thus the open sets in $\hat{C}$ which contain the point at infinity are those sets whose complements are compact (Fig. 22).

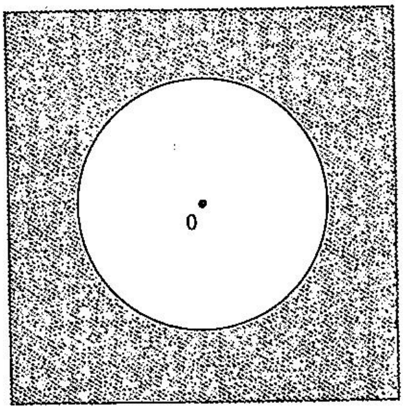


Fig. 21

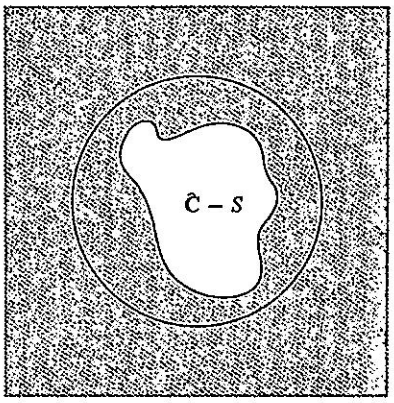


Fig. 22

We could pursue these topological ideas. (If you are interested, try Problems 5 and 6 in Section 13.6.) But we are more interested in functions on $\hat{C}$. After all, the reason we gave for considering the notion of a point at infinity amounted to this: that we can simplify the discussion of bilinear transformations by imagining them as transformations of the *extended* complex plane onto itself.

A **function** $\hat{f}$ which maps the extended complex plane to itself is a set of ordered pairs (z, w) *where* z *and* w *lie in* $\hat{C}$, with the usual rule that if (z, w) and (z, w') lie in the set then $w = w'$. Our reason for being pedantic about the definition is to point out that we can write down some perfectly reasonable statements about functions on $\hat{C}$ which look rather surprising in usual notation: for example

$$\hat{f}(1 + 2i) = \infty, \quad \text{and} \quad \hat{f}(\infty) = 1 - i.$$

Next we look at limits and continuity for functions defined on $\hat{C}$.

Definition

A function $\hat{f} : \hat{C} \longrightarrow \hat{C}$ has **limit** l (where $l \in \hat{C}$) **near** a (where $a \in \hat{C}$) if, given any neighbourhood N_l of l, there is a punctured neighbourhood P_a of a such that $\hat{f}(z) \in N_l$ whenever $z \in P_a$.

(If you think about this for a moment you will see that if l and a both lie in C, this is the usual definition. If $l \in C$ but $a = \infty$ then the new definition is equivalent to the one we gave for $\lim_{z \to \infty} f(z) = l$, whereas, if $l = \infty$ but $a \in C$ it is equivalent to the definition we gave for $\lim_{z \to a} f(z) = \infty$. The one new case is $\lim_{z \to \infty} f(z) = \infty$. We have to rephrase the definition in terms of neighbourhoods and punctured neighbourhoods because we cannot describe a neighbourhood of ∞ in the same way as we normally describe a neighbourhood of a point $a \in C$: that is, $\{z : |z - a| < \varepsilon\}$. A punctured neighbourhood of ∞ is just a neighbourhood of ∞ with ∞ itself omitted, in other words $\{z : z \in C \text{ and } |z| > r\}$.)

Definition

A function $\hat{f} : \hat{C} \longrightarrow \hat{C}$ is **continuous at** $a \in \hat{C}$ if $\lim_{z \to a} \hat{f}(z) = \hat{f}(a)$.

The natural question to ask is this. Suppose we are given a continuous function $f:\mathbf{C} \longrightarrow \mathbf{C}$; can we extend f to a continuous function $\hat{f}:\hat{\mathbf{C}} \longrightarrow \hat{\mathbf{C}}$? In other words, can we find a continuous function $\hat{f}$ on the extended complex plane such that $\hat{f}(z) = f(z)$ when $z \in \mathbf{C}$? Well, if $\lim\limits_{z\to\infty} f(z) = l$, where $l \in \mathbf{C}$, we can define $\hat{f}$ as follows:

$$\hat{f}(z) = \begin{cases} f(z), & z \in \mathbf{C} \\ l, & z = \infty. \end{cases}$$

The function $\hat{f}$ is called a **continuous extension** of f to $\hat{\mathbf{C}}$. Now $\lim\limits_{z\to\infty} f(z) = \lim\limits_{z\to 0} f(1/z)$, and if f is analytic, our condition that $\lim\limits_{z\to\infty} f(z) \in \hat{\mathbf{C}}$ amounts to saying that the singularity of the function $z \longrightarrow f(1/z)$ at 0 is either removable (if $\lim\limits_{z\to\infty} f(z) \in \mathbf{C}$) or a pole (if $\lim\limits_{z\to\infty} f(z) = \infty$). (See *Unit 8*, Corollary to Theorem 1, Section 8.1, and Theorem 4, Section 8.3.) In accordance with our general rule about interpreting the point at infinity as "the image of 0 under the function $z \longrightarrow 1/z$", we call the singularity of $z \longrightarrow f(1/z)$ at 0 the **singularity of f at ∞**. So if f is entire, and has either a removable singularity or a pole at ∞ (in other words, does not have an essential singularity at ∞), then it can be extended to a continuous function on $\hat{\mathbf{C}}$. More generally, a function f, which has a number of poles in $\mathbf{C}$ (and no other singularities) and a removable singularity or a pole at ∞, can be extended to a continuous function $\hat{f}$ on $\hat{\mathbf{C}}$, where $\hat{f}(z) = \infty$ whenever z is a singularity of f.

A bilinear transformation $f(z) = \dfrac{az + b}{cz + d}$ satisfies these conditions, and so extends to a function $\hat{f}$ on $\hat{\mathbf{C}}$. Notice that $\hat{f}(-d/c) = \infty$ and $\hat{f}(\infty) = a/c$, if $c \neq 0$; whereas if $c = 0, \hat{f}(\infty) = \infty$. Since f is one-one, so is $\hat{f}$.

In fact, this last property characterizes bilinear functions.

Theorem 6

Suppose that the function f is analytic on $\mathbf{C}$ except for a number of poles, and has a removable singularity or pole at ∞; suppose also that $\hat{f}$, the continuous extension of f to $\hat{\mathbf{C}}$, is one-one from $\hat{\mathbf{C}}$ onto $\hat{\mathbf{C}}$. Then f is a bilinear transformation.

Proof

We note first of all that since $\hat{f}$ is one-one, the function f can have at most one pole in $\mathbf{C}$. There are two possibilities: either there are points α and β in $\mathbf{C}$ such that $\hat{f}(\alpha) = \infty$ and $\hat{f}(\infty) = \beta$, or $\hat{f}(\infty) = \infty$. In the first case, f has a pole at α, and in the second it has no singularity in $\mathbf{C}$.

Suppose first that $\hat{f}(\alpha) = \infty$ and $\hat{f}(\infty) = \beta$, where $\alpha, \beta \in \mathbf{C}$. Then f has a pole at α; suppose that its order is r, where $r \geqslant 1$. We may write

$$f(z) = \frac{a_{-r}}{(z-\alpha)^r} + \frac{a_{-r+1}}{(z-\alpha)^{r-1}} + \cdots + \frac{a_{-1}}{(z-\alpha)} + g(z), \quad z \neq \alpha, z \in \mathbf{C},$$

where $a_{-r} \neq 0$ and g is an entire function. Now $\lim\limits_{z\to\infty} f(z) = \hat{f}(\infty) = \beta$, and $\lim\limits_{z\to\infty} \dfrac{1}{(z-\alpha)^s} = 0$ for $s = 1, 2, \ldots, r$. It follows that $\lim\limits_{z\to\infty} g(z) = \beta$. This means that g is bounded. For there is some N such that $|g(z) - \beta| < 1$ when $|z| > N$, and so $|g(z)| < 1 + |\beta|$ when $|z| > N$. But g is continuous, and the set $\{z : |z| \leqslant N\}$ is compact, so g is bounded on this set (by the Boundedness Theorem of *Unit 2*). Suppose that k is a bound for g on $\{z : |z| \leqslant N\}$. Then $|g(z)| \leqslant \max\{k, 1 + |\beta|\}$ for all z in $\mathbf{C}$, and g is bounded. Thus by Liouville's Theorem (*Unit 5*) g is constant, and since $\lim\limits_{z\to\infty} g(z) = \beta$, we have $g(z) = \beta$ for all z in $\mathbf{C}$. So

$$f(z) = \frac{a_{-r}}{(z-\alpha)^r} + \frac{a_{-r+1}}{(z-\alpha)^{r-1}} + \cdots + \frac{a_{-1}}{(z-\alpha)} + \beta, \quad z \neq \alpha, z \in \mathbf{C}.$$

Now f must be one-one from $\mathbf{C} - \{\alpha\}$ to $\mathbf{C} - \{\beta\}$, and so its derivative cannot be 0. The derivative of f is

$$f'(z) = -\frac{ra_{-r}}{(z-\alpha)^{r+1}} - \frac{(r-1)a_{-r+1}}{(z-\alpha)^r} - \cdots - \frac{a_{-1}}{(z-\alpha)^2}, \quad z \neq \alpha, z \in \mathbf{C},$$

and so

$$(z-\alpha)^{r+1}f'(z) = -ra_{-r} - (r-1)a_{-r+1}(z-\alpha) - \cdots - a_{-1}(z-\alpha)^{r-1}.$$

Now if $r > 1$, the right-hand side is a non-constant polynomial, say $p(z)$; and by the Fundamental Theorem of Algebra there is some $\gamma \in \mathbf{C}$ such that $p(\gamma) = 0$. Moreover, $\gamma \neq \alpha$, since $a_{-r} \neq 0$. But then $f'(\gamma) = 0$. This possibility is excluded by the fact that f is one-one. Thus $r = 1, a_{-1} \neq 0$, and

$$f(z) = \frac{a_{-1}}{z-\alpha} + \beta,$$

which is bilinear.

Suppose on the other hand, that $\hat{f}(\infty) = \infty$. Let $\phi(z) = 1/f(z)$. Then $\hat{\phi}$ is a one-one transformation of $\hat{\mathbf{C}}$ onto itself; $\hat{\phi}(\infty) = 0$ and if α is the complex number such that $f(\alpha) = 0$, then $\hat{\phi}(\alpha) = \infty$. Thus ϕ belongs to the class of functions we have already considered, and so is bilinear:

$$\phi(z) = \frac{a_{-1}}{z-\alpha},$$

and so f is also bilinear,

$$f(z) = \frac{1}{a_{-1}} \cdot (z-\alpha). \quad \blacksquare$$

Notice that a one-one transformation of $\mathbf{C}$ onto itself which leaves ∞ fixed is the extension of a similarity, $z \longrightarrow az + b$. We use this result in Section 13.7.

Let us summarize what we have done so far. We have pointed out that discussion of the properties of bilinear transformations can be simplified if we have a notion of a point at infinity; we have described how the complex plane can be extended by the addition of such a point and given a definition of the neighbourhood of ∞. We have shown how suitable functions on $\mathbf{C}$ can be extended to functions on $\hat{\mathbf{C}}$. Also we have proved a theorem which shows that bilinear transformations are characterized by the fact that they extend to one-one, continuous mappings of $\hat{\mathbf{C}}$ onto itself.

This is all very well; but our approach to the discussion of the point at infinity has been very abstract. It would obviously be helpful if we could visualize the extended complex plane in some way. We can in fact visualize the extended complex plane—as a sphere, related to the plane $\mathbf{C}$ by *stereographic projection*. This we now explain.

Imagine a sphere, with a point on it distinguished—the north pole, say. The tangent plane at the point diametrically opposite the north pole, that is, the south pole, is to be regarded as the complex plane (not yet extended). Let the sphere touch the complex plane at the origin. To each point in the plane we make correspond a point of the sphere, as follows. The point on the sphere corresponding to the point P on the plane is the point other than the north pole at which the (straight) line joining P to the north pole intersects the sphere. See Fig. 23 on page 49. The mapping which takes a point of the plane to the corresponding point of the sphere is a **stereographic projection**.*

---

* Another stereographic projection is obtained by considering the plane through the equator of the sphere to be the complex plane.

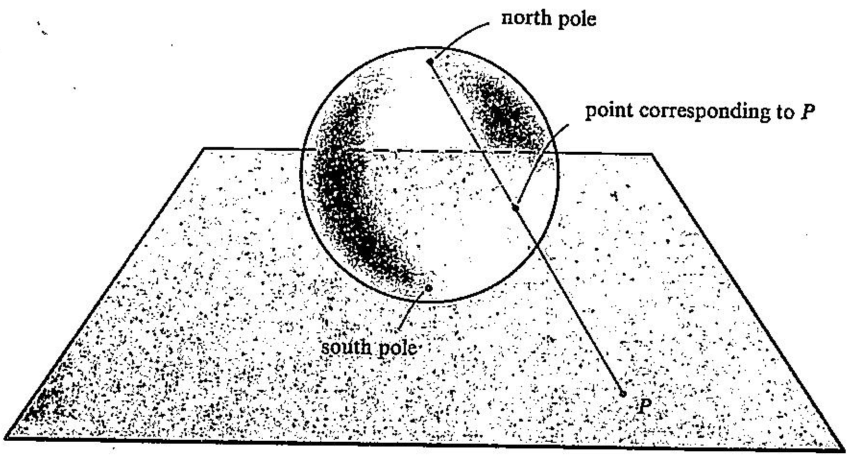


Fig. 23

Now every point on the sphere, *with the exception of the north pole*, is the image of exactly one point on the plane. Moreover, any neighbourhood of a point P in the plane is mapped to a nice set on the sphere, containing the stereographic image of P (Fig. 24). You should be able to convince yourself that this stereographic projection is a continuous mapping from the plane to the sphere with its north pole missing. (We are using the word "continuous" in a rather loose way here. We would like you to recognize, however intuitively, that topological properties of the plane are preserved by stereographic projection.*)

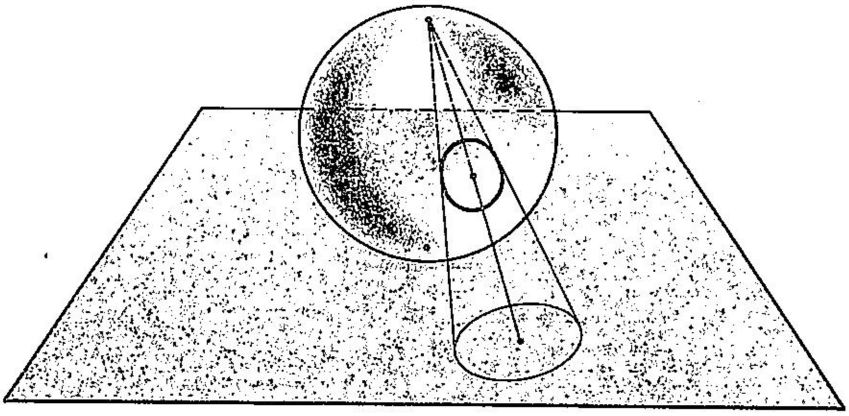


Fig. 24

The whole sphere, then, can be thought of as representing the complex plane with one extra point added, namely the north pole. Moreover, the stereographic image of the outside of a circle centre 0 in the plane—which is a punctured neighbourhood of ∞—is a small spherical cap around the north pole (Fig. 25).

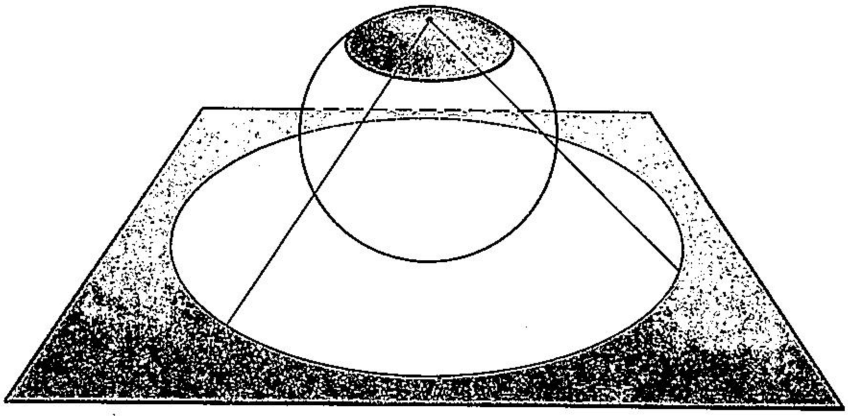


Fig. 25

---

*In fact, stereographic projection is a homeomorphism, as former students of M202 may recognize.

This means that the whole sphere gives us a model of the extended complex plane, both as a set of points, and topologically, in that neighbourhoods of points in the extended complex plane corresponding to something sensible. The north pole represents the point at infinity.

The sphere is a model of the extended complex plane, which faithfully reflects its topological properties. Does it reflect the plane's geometrical properties? Evidently it does not do so completely, because distances are distorted. A pair of points close together and near the north pole of the sphere correspond to a pair of points in the plane which may be very far apart; but a pair of points close together and near the south pole correspond to a pair of points in the plane which are also close together. On the other hand, angles are preserved by stereographic projection. This may be seen as follows.

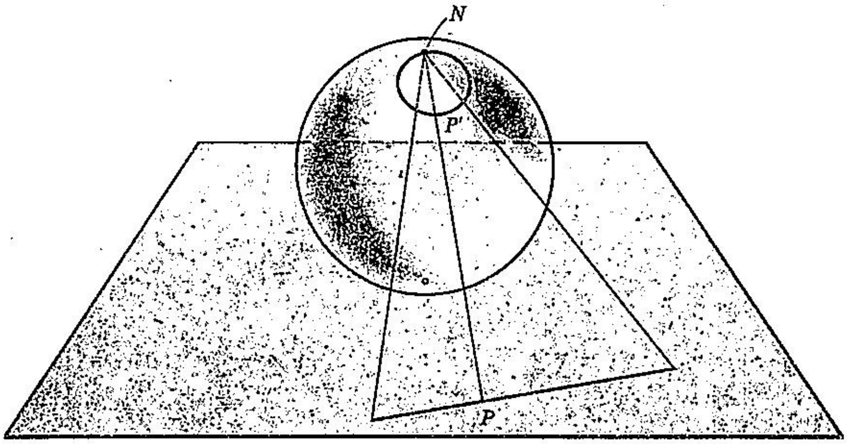


Fig. 26

Let P be a point of the plane, P' its stereographic image, and N the north pole of the sphere. Consider a line through P in the plane. Its stereographic image is the curve in which the plane through N and P containing the line cuts the sphere (Fig. 26). A plane cuts a sphere in a circle; so the image of the line is a circle, which passes through P' and N. (A line is a circle which passes through the point at infinity!) Now consider two directions at P, represented by rays. You may imagine them as the tangents at P to two curves which intersect at P. Their stereographic images will be two circular arcs joining P' and N.

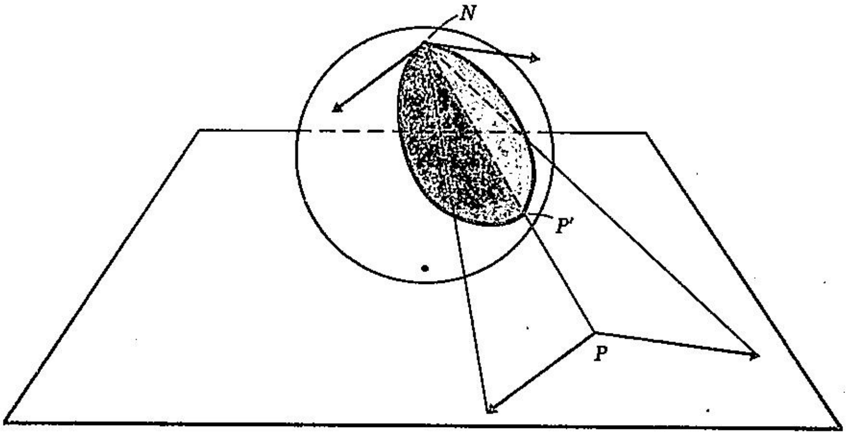


Fig. 27

We have to show that the angle between these circular arcs at P'—that is, the angle between their tangents at P'—is the same as the angle between the directions at P (Fig. 27). Now the angle between the circular arcs at P' is the same as the angle between them where they meet again at N. But the tangent plane to the sphere at N is parallel to its tangent plane at the south pole, and the tangents to the circular arcs at N are parallel to the rays at P. So the angle between the circular arcs at N is equal to the angle at P with which we started.

As a further example of the relation between the geometry of the sphere and of the plane, we shall determine which transformations of the plane correspond to the rotations of the sphere, and use our answer to show that the image under stereographic projection of any circle or line in the plane is a circle on the sphere.

A transformation of the sphere defines, in an obvious way, a transformation of the extended complex plane. A rotation of the sphere (through some angle about one of its diameters) has the following properties: it is one-one, conformal (angle preserving), and maps the end-points of any diameter of the sphere to a pair of points which are also diametrically opposite. Since stereographic projection preserves angles, the corresponding transformation of the plane is also conformal, and thus analytic (by Problem 2 of Section 13.2). A rotation of the sphere thus defines a one-one transformation of the extended complex plane which is analytic, and is therefore a bilinear transformation by Theorem 6 above. We must investigate the limitations imposed by the condition that the transformation maps diametrically opposite points to diametrically opposite points. It turns out that if z is a point of the plane, the point corresponding to the opposite end of the diameter of the sphere through the stereographic image of z is $-1/\bar{z}$ (see Problem 3(c) of Section 13.6). Thus if the bilinear transformation $z \longrightarrow \dfrac{az+b}{cz+d}$ is to correspond to a rotation of the sphere, then the point

$$\frac{a(-1/\bar{z})+b}{c(-1/\bar{z})+d}$$

must correspond to the point diametrically opposite the stereographic image of $\dfrac{az+b}{cz+d}$. Thus we must have

$$\frac{a(-1/\bar{z})+b}{c(-1/\bar{z})+d} = -\frac{\bar{c}\bar{z}+\bar{d}}{\bar{a}\bar{z}+\bar{b}} \quad \text{for all } z.$$

This means that

$$(-a+b\bar{z})(\bar{a}\bar{z}+\bar{b}) + (-c+d\bar{z})(\bar{c}\bar{z}+\bar{d}) = 0 \quad \text{for all } z,$$

that is

$$(\bar{a}b+\bar{c}d)\bar{z}^2 - (a\bar{a}-b\bar{b}+c\bar{c}-d\bar{d})\bar{z} - (a\bar{b}+c\bar{d}) = 0 \quad \text{for all } z.$$

It must therefore be the case that

$$\bar{a}b+\bar{c}d = 0 \tag{1}$$

and

$$a\bar{a}-b\bar{b}+c\bar{c}-d\bar{d} = 0. \tag{2}$$

We may also assume that

$$ad-bc = 1 \quad \text{(see Problem 1(vi) of Section 13.4).}$$

Let us set $d = \lambda\bar{a}$. This merely serves to define λ, if $a \neq 0$. If $a = 0$, on the other hand, then $\bar{c}d = 0$, and so $d = 0$, since $ad - bc \neq 0$; and then the equation $d = \lambda\bar{a}$ is certainly satisfied whatever λ may be. Substituting $d = \lambda\bar{a}$ in Equation (1), we obtain

$$\bar{a}(b+\lambda\bar{c}) = 0.$$

Thus $b + \lambda\bar{c} = 0$, if $a \neq 0$; whereas, if $a = 0$, we can use the equation $b + \lambda\bar{c} = 0$ to define λ. Substituting $d = \lambda\bar{a}$, $b = -\lambda\bar{c}$ in Equation (2), we obtain

$$(1-|\lambda|^2)(|a|^2+|c|^2) = 0.$$

We cannot have $a = c = 0$, so $|\lambda| = 1$. On the other hand, substituting $d = \lambda\bar{a}$, $b = -\lambda\bar{c}$ in the equation $ad - bc = 1$, we have

$$\lambda(|a|^2+|c|^2) = 1.$$

Thus λ is real, positive, and of unit modulus, and so $\lambda = 1$. The bilinear transformation therefore has the form

$$z \longrightarrow \frac{az + b}{-\bar{b}z + \bar{a}}, \quad \text{where } |a|^2 + |b|^2 = 1.$$

(The condition $|a|^2 + |b|^2 = 1$ is not significant, since it reflects the fact that the coefficients of a bilinear transformation are determined only up to a factor.) Conversely, a bilinear transformation of the form $z \longrightarrow \frac{az + b}{-\bar{b}z + \bar{a}}$ defines a one-one, conformal transformation of the sphere, which maps opposite ends of a diameter to opposite ends of a diameter, and is therefore a rotation.

Notice the similarity between the rotations of the sphere, which are of the form $z \longrightarrow \frac{az + b}{-\bar{b}z + \bar{a}}$, and the bilinear transformations mapping the unit circle onto itself, which are of the form $z \longrightarrow \frac{az + b}{\bar{b}z + \bar{a}}$. Notice the similarity, but do not be misled by it. The extra minus sign makes a crucial difference.

Let us turn our attention now to the question: what are the stereographic images of circles in the plane? (We already know that the stereographic image of a line in the plane is a circle through the north pole of the sphere.) It is geometrically very clear what the answer is for a particular class of circles, those with centre at the origin. The stereographic image of such a circle is a circle on the sphere, which is the intersection of a horizontal plane with the sphere (Fig. 28).

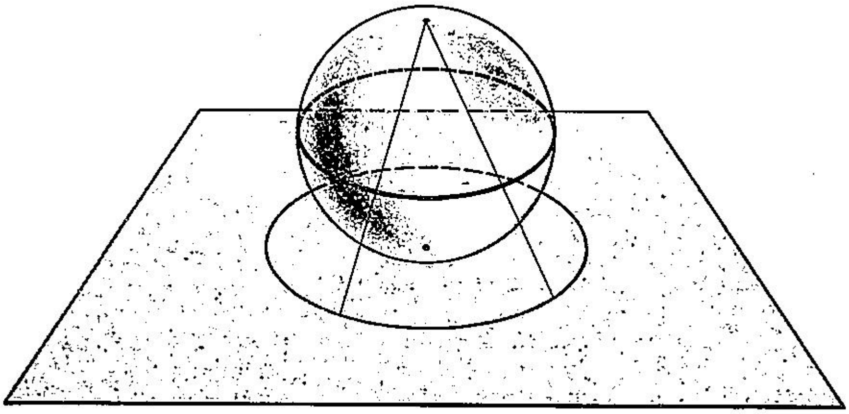


Fig. 28

Any other circle may be transformed, by means of a bilinear transformation of the form $z \longrightarrow \frac{az + b}{-\bar{b}z + \bar{a}}$, to a circle centre the origin. (See Problem 4, Section 13.6.) Its stereographic image may be constructed in stages. First carry out the bilinear transformation which transforms it to a circle centre 0; then stereographically project; then carry out the rotation of the sphere which is the inverse of the rotation corresponding to the initial bilinear transformation. At each stage the image curve is a circle: the bilinear transformation maps circles to circles; the stereographic image of a circle centre 0 is a circle; rotations of the sphere map circles to circles. So the stereographic image of any circle or line in the plane is a circle on the sphere.

Summary

The complex plane may be extended by the addition of a point at infinity, together with a notion of a neighbourhood of ∞. One advantage of this process is that it enables us to dispense with all the special cases when discussing bilinear transformations. For example, it becomes sensible, and correct, to say that a bilinear transformation maps circles to circles. Moreover, we are able to characterize the bilinear transformations as those transformations which may be extended to one-one continuous transformations of the extended complex plane onto itself.

The extended complex plane may be visualized as the surface of a sphere, related to the plane by stereographic projection. The north pole of the sphere corresponds to the point at infinity. This gives us a conformal model of the extended plane, in which circles and lines in the extended plane correspond to circles on the sphere.

Self-Assessment Questions

1. Which of the following sets contains a punctured neighbourhood of ∞?
   (i) The closed disc $\{z : |z - 1| \leqslant 1\}$.
   (ii) The set $\{z : |z - 1| \geqslant 1\}$, which is the outside of an open disc.
   (iii) The right-hand half-plane $\{z : \operatorname{Re} z > 0\}$.

2. Find $\hat{f}(\infty)$, if $\hat{f}$ is the continuous extension to $\hat{\mathbb{C}}$ of the function $f(z) = \dfrac{z+i}{iz+1}$. Find also $\hat{f}^{-1}(\infty)$.

3. Describe as completely as you can the curves in the plane whose stereographic images are
   (i) lines of latitude (lines parallel to the equator);
   (ii) lines of longitude (lines through the poles).

4. Which of the following transformations correspond to rotations of the sphere?
   (i) $z \longrightarrow e^{i\theta}z$. (ii) $z \longrightarrow \dfrac{1}{z}$. (iii) $z \longrightarrow \dfrac{z-i}{iz+1}$.

5. Classify the singularity at ∞ for each of the following functions.
   (i) $f(z) = z^3$. (ii) $f(z) = e^z$. (iii) $f(z) = \dfrac{1}{z}$.

Solutions

1. The only set of the three which contains a punctured neighbourhood of ∞ is the second one, $\{z : |z - 1| \geqslant 1\}$.

2. $\hat{f}(\infty) = \lim_{z \to \infty} \dfrac{z+i}{iz+1} = \dfrac{1}{i} = -i$.

   $\hat{f}^{-1}(\infty)$ is the point which is mapped to ∞ by $\hat{f}$, which is i.

3. (i) Lines of latitude are stereographic images of circles centre 0.
   (ii) Lines of longitude are stereographic images of straight lines through 0.

4. (i) $z \longrightarrow e^{i\theta}z$ corresponds to a rotation, in which $a = e^{i\theta/2}$ and $b = 0$.
   (ii) $z \longrightarrow 1/z$ corresponds to a rotation, in which $a = 0$ and $b = i$.
   (iii) $z \longrightarrow \dfrac{z-i}{iz+1}$ does not correspond to a rotation.

5. (i) Pole of order 3. (ii) Essential singularity. (iii) Removable singularity.

13.6 PROBLEMS

1. (a) Show that $\lim_{z \to a} f(z) = \infty$ if and only if $\lim_{z \to a} \frac{1}{f}(z) = 0$.

   (b) Suppose that the function f has a number of poles in $\mathbf{C}$, and no other singularities, and a removable singularity or pole at ∞. Show that if $\phi = \frac{1}{f}$, then ϕ has poles as its only non-removable singularities.

   (c) Let $\hat{f}$ and $\hat{\phi}$ be the continuous extensions of f and ϕ to $\hat{\mathbf{C}}$. Show that if $\hat{f}(a) = \infty$ then $\hat{\phi}(a) = 0$, and if $\hat{f}(b) = 0$ then $\hat{\phi}(b) = \infty$.

2. If a function f is analytic on a neighbourhood of ∞, say $\{z : |z| > r\}$, then we may expand it in a Laurent series which converges for $|z| > r$:

$$f(z) = \cdots + \frac{a_{-n}}{z^n} + \frac{a_{-n+1}}{z^{n-1}} + \cdots + \frac{a_{-1}}{z} + a_0 + a_1 z + \cdots + a_n z^n$$
$$+ a_{n+1} z^{n+1} + \cdots.$$

   Explain how we can recognize from this Laurent series what kind of singularity f has at ∞.

3. In this question we shall discuss stereographic projection more quantitatively than in the text. You will need to do some coordinate geometry to tackle it successfully.

   (a) The equation of a sphere in $\mathbf{R}^3$ (coordinates X, Y and Z) whose centre is at $(0, 0, r)$ and whose radius is r is

$$X^2 + Y^2 + (Z - r)^2 = r^2.$$

   The north pole of the sphere is at $(0, 0, 2r)$. The tangent plane at the south pole, $(0, 0, 0)$, is the XY-plane (Fig. 29). If $(x, y, 0)$ is a point in the XY-plane, corresponding to the complex number $z = x + iy$, then any point on the line from this point to the north pole has the form $(\lambda x, \lambda y, (1 - \lambda)2r)$ for some (real) number λ. Find for which values of λ this point lies on the sphere, and so find the coordinates of the stereographic image of z.

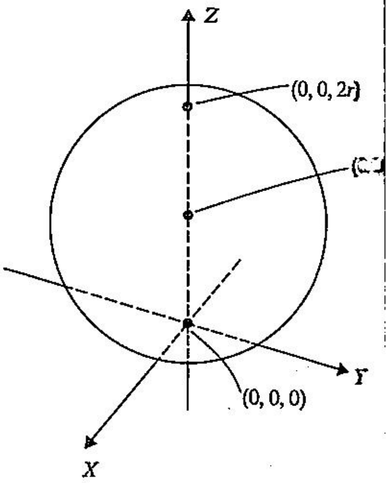


Fig. 29

   (b) The intersection of the sphere with a horizontal plane $Z = k$, where $0 < k < 2r$, is a circle parallel to the equator. Show that such a circle is the stereographic image of a circle centre 0 in the complex plane, and deduce that if $r = \frac{1}{2}$, then the unit circle in the complex plane stereographically projects onto the equator of the sphere. (This seems quite a good reason for choosing $r = \frac{1}{2}$.)

   (c) Show that the point in the complex plane whose stereographic image is the opposite end of the diameter of the sphere through the stereographic image of z is the point $-1/\bar{z}$.

4. Show that given any circle $Az\bar{z} + \bar{B}z + B\bar{z} + C = 0$, there is a bilinear transformation of the form $z \longrightarrow \dfrac{az + b}{-\bar{b}z + \bar{a}}$ which maps the circle to a circle centre 0.

5. Show that the extended complex plane is compact. That is, show that if to each point z of $\hat{\mathbf{C}}$ is assigned a neighbourhood N_z, then no matter how the neighbourhoods are chosen, there will be a finite number of points $z_1, z_2, \ldots, z_n$ such that

$$N_{z_1} \cup N_{z_2} \cup \cdots \cup N_{z_n} = \hat{\mathbf{C}}.$$

   (Hint: One of the points $z_1, z_2, \ldots, z_n$ is ∞.)

6. A simple-closed polygon in $\hat{\mathbf{C}}$—for example, a straight line—divides $\hat{\mathbf{C}}$ into two regions, though it is not necessarily true that either of them is bounded. We adapt the definition of simple-connectedness from *Unit 9*, Section 9.3, to cover regions in $\hat{\mathbf{C}}$, as follows. A region R in $\hat{\mathbf{C}}$ is called **simply-connected** in $\hat{\mathbf{C}}$ if whenever Γ is a simple polygon in R then one of the two regions making up the complement of Γ in $\hat{\mathbf{C}}$ is contained in R.

   For each of the following regions R, decide whether or not it is simply-connected in $\hat{\mathbf{C}}$.

   (i) $R = \mathbf{C} = \hat{\mathbf{C}} - \{\infty\}$ (the complex plane).

   (ii) $R = \hat{\mathbf{C}} - \{0\}$ (the extended plane punctured at 0).

   (iii) $R = \hat{\mathbf{C}} - \{0 \cup \infty\}$ (the plane punctured at 0).

   (iv) $R = \{z : \operatorname{Im} z > 0, z \neq \infty\}$ (the upper half-plane).

   (v) $R = \{z : |z| > r\} \cup \{\infty\}$ (a neighbourhood of ∞).

Solutions

1. (a) Suppose that $\lim_{z \to a} f(z) = \infty$. Then given $\varepsilon > 0$ there is $\delta > 0$ such that $|f(z)| > 1/\varepsilon$ when $0 < |z - a| < \delta$. Thus given $\varepsilon > 0$, there is $\delta > 0$ such that $\left|\frac{1}{f}(z)\right| < \varepsilon$ whenever $0 < |z - a| < \delta$, and so $\lim_{z \to a} \frac{1}{f}(z) = 0$.

   Suppose that $\lim_{z \to a} \frac{1}{f}(z) = 0$. Then given $N > 0$, there is $\delta > 0$ such that $\left|\frac{1}{f}(z)\right| < \frac{1}{N}$ when $0 < |z - a| < \delta$. Thus given $N > O$, there is $\delta > 0$ such that $|f(z)| > N$ when $0 < |z - a| < \delta$, and so $\lim_{z \to a} f(z) = \infty$.

   (b) If f has a pole at a then $\lim_{z \to a} \phi(z) = 0$, so ϕ has a removable singularity at a. The only non-removable singularities of ϕ occur at the zeros of f, and are therefore poles (see *Unit 8*, Corollary to Theorem 3, page 15).

   (c) If $\hat{f}(a) = \infty$, then $\lim_{z \to a} f(z) = \infty$, and so $\lim_{z \to a} \phi(z) = 0$, and $\hat{\phi}(a) = 0$. If $\hat{f}(b) = 0$, then b is a pole of ϕ, and so $\hat{\phi}(b) = \infty$.

2. Let $\phi(z) = f(1/z)$. Then

$$\phi(z) = \cdots + a_{-n}z^n + a_{-n+1}z^{n-1} + \cdots + a_{-1}z + a_0 + \frac{a_1}{z} + \cdots$$
$$+ \frac{a_n}{z^n} + \frac{a_{n+1}}{z^{n+1}} + \cdots;$$

   This Laurent series for ϕ converges for $0 < |z| < 1/r$.

   Thus ϕ has:
   a removable singularity at 0 if $a_n = 0$ for all $n > 0$;
   a pole of order k at 0 if $a_n = 0$ for all $n > k > 0$, but $a_k \neq 0$;
   an essential singularity at 0 if there are infinitely many non-zero coefficients a_n with $n > 0$.

   Thus f has:
   a removable singularity at ∞ if its Laurent series contain no positive powers of z;
   a pole of order k at ∞ if its Laurent series contains no positive powers of z of order higher than k;
   an essential singularity at ∞ if its Laurent series contains arbitrarily large positive powers of z.

3. (a) Notice first of all that the equation of the sphere may be rewritten as

$$X^2 + Y^2 + Z^2 - 2rZ = 0,$$

   that is

$$X^2 + Y^2 + Z(Z - 2r) = 0.$$

In order that the point $(\lambda x, \lambda y, (1-\lambda)2r)$ lie on the sphere, we must have

$$\lambda^2 x^2 + \lambda^2 y^2 + 2(1-\lambda)r\cdot(-\lambda\cdot 2r) = 0,$$

that is

$$\lambda^2(x^2+y^2+4r^2) - 4\lambda r^2 = 0.$$

Thus either $\lambda = 0$ (which gives the north pole) or

$$\lambda = \frac{4r^2}{4r^2+|z|^2} \quad \text{(where } z = x+iy\text{)}.$$

Thus the stereographic image of $z = x+iy$ is the point $(\lambda x, \lambda y, (1-\lambda)2r)$ corresponding to this value of λ, which is

$$\left(\frac{4r^2x}{4r^2+|z|^2}, \frac{4r^2y}{4r^2+|z|^2}, \frac{2r|z|^2}{4r^2+|z|^2}\right).$$

(b) The points z in the complex plane whose stereographic images lie on the intersection of the sphere with the plane $Z = k$ are those for which

$$\frac{2r|z|^2}{4r^2+|z|^2} = k,$$

that is,

$$|z|^2 = \frac{4r^2k}{2r-k}.$$

They therefore lie on a circle centre the origin.

The stereographic image of the unit circle is the intersection of the sphere with that horizontal plane $Z = k$ for which $\dfrac{4r^2k}{2r-k} = 1$. Then $k = \dfrac{2r}{1+4r^2}$. This will be the equator of the sphere provided $k = r$. But if $\dfrac{2r}{1+4r^2} = r$, then $r = \frac{1}{2}$.

(c) With the choice $r = \frac{1}{2}$, the stereographic image of $z = x+iy$ is

$$\left(\frac{x}{1+|z|^2}, \frac{y}{1+|z|^2}, \frac{|z|^2}{1+|z|^2}\right).$$

The point on the sphere diametrically opposite this point is

$$\left(-\frac{x}{1+|z|^2}, -\frac{y}{1+|z|^2}, 1-\frac{|z|^2}{1+|z|^2}\right),$$

that is,

$$\left(-\frac{x}{1+|z|^2}, -\frac{y}{1+|z|^2}, \frac{1}{1+|z|^2}\right).$$

But this is just the stereographic image of the point

$$-\frac{1}{\bar{z}} = -\frac{x}{|z|^2} - i\frac{y}{|z|^2}.$$

4. Suppose the equation of the circle is

$$Az\bar{z} + \bar{B}z + B\bar{z} + C = 0.$$

Let $w = \dfrac{az+b}{-\bar{b}z+\bar{a}}$; then $z = \dfrac{\bar{a}w-b}{\bar{b}w+a}$, and so the image of the circle is

$$A(\bar{a}w-b)(a\bar{w}-\bar{b}) + \bar{B}(\bar{a}w-b)(b\bar{w}+\bar{a})$$
$$+B(a\bar{w}-\bar{b})(\bar{b}w+a) + C(\bar{b}w+a)(b\bar{w}+\bar{a}) = 0.$$

This will have centre 0 if the coefficient of $\bar{w}$ is zero (see page 25), that is

$$-abA - b^2\bar{B} + a^2B + abC = 0.$$

If $b \neq 0$, this may be rewritten as a quadratic equation in a/b:

$$B\frac{a^2}{b^2} + (C-A)\frac{a}{b} - \bar{B} = 0,$$

and so values for a and b certainly exist.

5. Consider the complement in $\hat{\mathbf{C}}$ of N_∞. It is a compact subset in $\mathbf{C}$—in fact a closed disc D centre the origin. We have associated with each point z in D a neighbourhood N_z (which is an open disc centre z). Since D is compact, there is a finite number of points $z_1, z_2, \ldots, z_m$ in D such that $D \subseteq N_{z_1} \cup N_{z_2} \cup \cdots \cup N_{z_m}$. But if $N_{z_1} \cup N_{z_2} \cup \cdots \cup N_{z_m}$ covers the complement of N_∞, then $N_{z_1} \cup N_{z_2} \cup \cdots \cup N_{z_m} \cup N_\infty$ covers $\hat{\mathbf{C}}$, that is,

$$N_{z_1} \cup N_{z_2} \cup \cdots \cup N_{z_m} \cup N_\infty = \hat{\mathbf{C}}.$$

6. (i) Simply-connected in $\hat{\mathbb{C}}$.
   (ii) Simply-connected in $\hat{\mathbb{C}}$.
   (iii) Not simply-connected in $\hat{\mathbb{C}}$.
   (iv) Simply-connected in $\hat{\mathbb{C}}$.
   (v) Simply-connected in $\hat{\mathbb{C}}$.

13.7 GLOBAL MAPPING THEOREMS

In this section we discuss some questions of a global nature about one-one analytic transformations.

* * * * * * * *

We begin by introducing a new term.

Definition

> Let R and S be regions. A function f defined on R is an **analytic equivalence of R with S** if f is analytic and one-one and $f(R) = S$. Two regions are **analytically equivalent** if there is an analytic equivalence of one with the other.

Notice that if f is an analytic equivalence of R with S then f^{-1} is an analytic equivalence of S with R.

The notion of analytic equivalence has important applications in the discussion of boundary value problems, as you will discover if you read *Unit 16, Boundary Value Problems* (though in that unit we are forced to take a rather more restrictive definition than the one above). Indeed the investigation of boundary value problems gave a strong impetus to the consideration of the sort of question we shall attempt to answer in this section. But these questions are interesting in their own right, and we shall make no further reference to boundary value problems here.

There are broadly speaking four kinds of question about analytic equivalences.

(i) Given a region R, what can we say about the regions analytically equivalent to R?

(ii) Given regions R and S, is there an analytic equivalence of R with S?

(iii) Given a region R, can we identify the set of analytic equivalences of R with itself?

(iv) Given analytically equivalent regions R and S, can we identify the set of analytic equivalences of R with S?

Question (iv) is answered once we know the answer to question (iii); since if f and g are analytic equivalences of R with S, then $g^{-1} \circ f$ is an analytic equivalence of R with itself.

Let us consider a relatively simple question of type (ii) first. Let R be the whole complex plane **C**, and S a bounded region. Is there an analytic equivalence of R with S? In other words, is there a one-one analytic function f whose domain is the whole of **C**, such that $|f(z)| < k$ for all z, where k is some given positive number? Any such function would be bounded and entire, and so by Liouville's Theorem it would be a constant. In other words, it would map each point of **C** to one and the same point, and so would be just as far from being one-one as is possible. So Liouville's Theorem, which up to now has appeared in a purely analytic guise, has an interesting geometrical consequence: the complex plane cannot be analytically equivalent to a bounded region.

We have answered a particular question of type (ii). Let us look at a type (i) question. What kind of region is analytically equivalent to **C**? Certainly **C** is analytically equivalent to itself. As it happens, **C** cannot be analytically equivalent to anything else. This we shall now prove, and at the same time answer a question of type (iii): what are the analytic equivalences of **C** with itself?

Theorem 7

Suppose that f is an analytic equivalence of $\mathbf{C}$ with a region S. Then $S = \mathbf{C}$ and f is a similarity.

In the proof we shall use the Casorati–Weierstrass Theorem (*Unit 8*, Theorem 5). It applies to a function g which has an essential singularity at ∞, and says that given any complex number w_0, any positive real number ε, and any punctured neighbourhood of ∞, there is a point z in the punctured neighbourhood of ∞ such that $|g(z) - w_0| < \varepsilon$.

Proof

Since f is defined on the whole of $\mathbf{C}$, it is defined and analytic on a punctured neighbourhood of ∞. On the face of it, its singularity at ∞ may be removable, essential, or a pole. We shall show first that it cannot be essential.

Suppose that the analytic equivalence f had an essential singularity at ∞. By the Open Mapping Theorem (page 11), the image of the unit disc $\{z : |z| < 1\}$ is an open set, and so contains a neighbourhood $\{w : |w - f(0)| < \varepsilon\}$. Let r be any real number greater than 1. Then $\{z : |z| > r\}$ is a punctured neighbourhood of ∞ disjoint from the unit disc. If f had an essential singularity at ∞, then by the Casorati–Weierstrass Theorem there would be a point z_0 such that $|z_0| > r$ and $|f(z_0) - f(0)| < \varepsilon$. But then $f(z_0)$ would lie in the image of the unit disc, so there would be some point z_1 such that $|z_1| < 1$ and $f(z_1) = f(z_0)$. Since $|z_0| > r > 1$ whereas $|z_1| < 1$, clearly $z_1 \neq z_0$. But this would contradict the fact that f is one-one. Thus f cannot have an essential singularity at ∞.

The singularity of f at ∞ cannot be removable either. If it were removable, then $\lim_{z \to \infty} f(z)$ would exist. Suppose $\lim_{z \to \infty} f(z) = l$ (where $l \in \mathbf{C}$, that is, $l \neq \infty$). Then there would be some positive real number r such that $|f(z) - l| < 1$ whenever $|z| > r$, and so $|f(z)| < 1 + |l|$ whenever $|z| > r$. The function f would be bounded on the compact set $\{z : |z| \leqslant r\}$, and so bounded everywhere; and thus entire. It would therefore be constant, by Liouville's Theorem. So f cannot have a removable singularity at ∞ either.

It follows that the singularity of f at ∞ is a pole, and so $\lim_{z \to \infty} f(z) = \infty$. The function f extends to a one-one continuous function $\hat{f}$ on the extended complex plane, such that $\hat{f}(\infty) = \infty$. But we showed that such a function must be a similarity, in Theorem 6 of Section 13.5. Moreover, if f is a similarity, then $f(\mathbf{C}) = \mathbf{C}$. So if f is an analytic equivalence of $\mathbf{C}$ with a region S, then $S = \mathbf{C}$ and f is a similarity. ■

We can extend this result to cover the question of the analytic equivalences of the region $\mathbf{C} - \{\alpha\}$, where α is some point in $\mathbf{C}$.

Theorem 8

Suppose that f is an analytic equivalence of $\mathbf{C} - \{\alpha\}$ with a region S. Then $S = \mathbf{C} - \{\beta\}$ for some point β, and f is a bilinear transformation.

Proof

The function f has a singularity at α, which, we may as well assume is not removable, since otherwise Theorem 7 would apply. The Casorati–Weierstrass Theorem may be used to show that this singularity cannot be essential, much as it was used in the proof of Theorem 7. Thus the singularity is a pole, and $\lim_{z \to \alpha} f(z) = \infty$.

Now the region S cannot be the whole of $\mathbf{C}$, since otherwise f^{-1} would be an analytic equivalence of $\mathbf{C}$ with $\mathbf{C} - \{\alpha\}$, and we know there cannot be such an analytic equivalence (by Theorem 7). There is therefore some point β not lying in S. Let $\phi(z) = \dfrac{1}{f(z) - \beta}$. Then ϕ is analytic on $\mathbf{C} - \{\alpha\}$, since we know that there is no z in $\mathbf{C} - \{\alpha\}$ such that $f(z) = \beta$. The singularity of ϕ at α is removable, since $\lim_{z \to \alpha} \phi(z) = 0$. The function obtained from ϕ by removing the singularity is entire. It is evidently one-one on $\mathbf{C} - \{\alpha\}$, and since $\phi(z) \neq 0$ for $z \neq \alpha$, it is actually one-one on $\mathbf{C}$. Thus by Theorem 7, ϕ is a similarity, say $\phi(z) = az + b$. Then $f(z) = \dfrac{1}{az + b} + \beta$, where $a\alpha + b = 0$; f is a bilinear transformation, and S is the image of $\mathbf{C} - \{\alpha\}$ under f, which is $\mathbf{C} - \{\beta\}$. ■

These results, together with Theorem 6 of Section 13.5, go a long way towards showing why bilinear transformations are so important in the theory of mapping properties of analytic functions.

We have dealt with analytic equivalences of C and of C with one point removed, and we have seen what little freedom of action there is in either case. Suppose we remove a bit more of the complex plane. What happens then? The situation changes dramatically, as the following celebrated theorem, due to Riemann, describes.

Theorem 9 (The Riemann Mapping Theorem)

Let R be a simply-connected region other than $\mathbf{C}$. Then R is analytically equivalent to the open unit disc $\{z : |z| < 1\}$.

The real force of this theorem is clearer if we observe that it implies that *any* two simply-connected regions, neither of which is C, are analytically equivalent. For if R and S are the regions in question, f is an analytic equivalence of R with the unit disc, and g an analytic equivalence of S with the unit disc, then $g^{-1} \circ f$ is an analytic equivalence of R with S.

(In *Unit 16* we state a weaker version of this result, in which R is assumed to have as boundary a simple-closed contour. Here we have given the theorem in full generality.)

The proof of this theorem is too difficult to be given in full in this course, but we can describe some of the geometric ideas behind it.

The first step is to use analytic equivalences to move R into a convenient starting position. In fact we may transform R into a region contained in the open unit disc. For example, if there is an interior point of $\mathbf{C} - R$, that is, a point α not in R such that there is a closed disc $|z - \alpha| \leqslant r$ which contains no point of R, then the transformation $z \longrightarrow \dfrac{r}{z - \alpha}$ maps R one-one onto a region R'. Since $|z - \alpha| > r$ for any z in R, it follows that $|w| = \left|\dfrac{r}{z - \alpha}\right| < 1$ for any w in R', in other words, that R' is contained in the open unit disc.

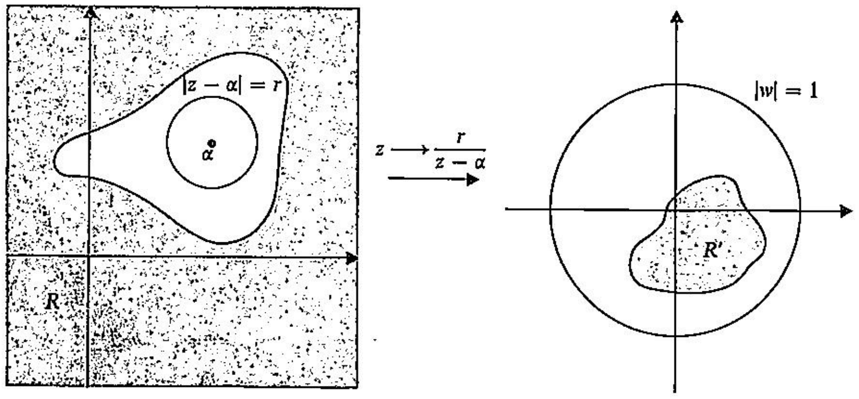


Fig. 30

Of course, there may not be an interior point of $\mathbf{C} - R$; but though the transformation required is more complicated, it is still true that R can be transformed into a region contained in the open unit disc. Moreover, it can be arranged that this region contains the point 0 (by carrying out a further, bilinear, transformation if necessary).

So we may as well assume that R is contained in the open unit disc, and so the problem becomes

> given a simply-connected region R contained in the open unit disc D, and containing 0, find an analytic equivalence of R with D.

Now we cannot explicitly give an analytic equivalence which solves this problem, in general, but we can construct a transformation ϕ which is analytic and has certain properties. The transformation of R onto D that we require is obtained, roughly speaking, by iterating ϕ. The following lemma sets out the properties of ϕ.

Lemma

Let R be a simply-connected region contained in the open unit disc D, let R contain 0. Then there exists a transformation ϕ which is analytic and is such that

(i) ϕ is an analytic equivalence of R with a region $\phi(R)$ contained in D;

(ii) for any z in R (other than 0), $|\phi(z)| > |z|$.

(Thus ϕ moves points of R towards the boundary of D.)

Proof

The boundary of R is a closed set. It is bounded (since it is contained in the disc $|z| < 1$); it is therefore compact. The function $z \longrightarrow |z|$ is continuous, and so attains its minimum value on the boundary of R (by the Minimum Principle, *Unit 9*). Let this minimum value be r. We first make a rotation, if necessary, so that the point at which this minimum is attained (or one of the points, if there are several) is moved to the point r. Since this rotation does not play a very important role in the remainder of the argument, we assume that it has been carried out, and ignore it. The situation is as shown in Fig. 31.

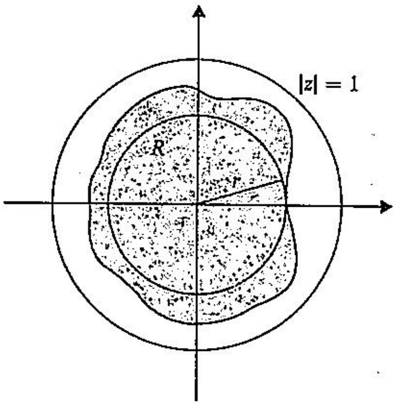


Fig. 31

Let $\theta_r(z) = \dfrac{z - r}{rz - 1}$, $0 < r < 1$. Then from Problem 2 of Section 13.4 we know that θ_r maps D onto D, and so θ_r maps R into D. It is of course, one-one, and analytic.

Now $\theta_r(r) = 0$, and r does not lie in R (it lies in the boundary of R). Thus $\theta_r(R)$ is a simply-connected region (see Problem 3 of Section 9.6 of *Unit 9*), which does not contain 0, so there is an analytic square root defined on it. Call this square root ρ. Then ρ is one-one on $\theta_r(R)$, and $\rho(\theta_r(R))$ is contained in D.

We define ϕ by

$$\phi = \theta_{\sqrt{r}} \circ \rho \circ \theta_r.$$

Since $0 < \sqrt{r} < 1$, the bilinear transformation $\theta_{\sqrt{r}}$ maps D onto D, and so maps $\rho(\theta_r(R))$ into D. Thus ϕ has the first of the two properties listed above.

We now show that it has the second property too. To do this, it is best to deal with ϕ^{-1} (since it involves squaring, which is simpler than taking square roots). Now $\theta_r^{-1} = \theta_r$ (by Problem 2 of Section 13.4), so

$$\phi^{-1} = \theta_r \circ \sigma \circ \theta_{\sqrt{r}}, \text{ where } \sigma(z) = z^2.$$

So for z in R,

$$\begin{aligned}
|z| &= |\phi^{-1}(\phi(z))| \\
&= \left|\theta_r\left(\left(\frac{\phi(z)-\sqrt{r}}{\sqrt{r}\phi(z)-1}\right)^2\right)\right| \\
&= \left|\frac{(\phi(z)-\sqrt{r})^2 - r(\sqrt{r}\phi(z)-1)^2}{r(\phi(z)-\sqrt{r})^2-(\sqrt{r}\phi(z)-1)^2}\right|, \quad \text{since } \theta_r(z) = \frac{z-r}{rz-1}, \\
&= \left|\frac{(1-r^2)(\phi(z))^2 - 2\sqrt{r}(1-r)\phi(z)}{2\sqrt{r}(1-r)\phi(z)-(1-r^2)}\right| \\
&= |\phi(z)|\cdot\left|\frac{(1+r)\phi(z)-2\sqrt{r}}{2\sqrt{r}\phi(z)-(1+r)}\right| \\
&= |\phi(z)|\cdot|\theta_s(\phi(z))|, \quad \text{where } s = \frac{2\sqrt{r}}{1+r} < 1.
\end{aligned}$$

But $|\phi(z)| < 1$, so $|\theta_s(\phi(z))| < 1$, and

$$|z| < |\phi(z)|,$$

which is what we wanted to prove. ■

We now apply ϕ again – or rather, since the distance of the boundary of $\phi(R)$ from 0 will be larger than r, we apply a new transformation which is nevertheless constructed exactly as ϕ is. Let us distinguish these various transformations by a subscript, so ϕ_n is the transformation used at the nth stage, with $\phi_1 = \phi$. Let $f_n = \phi_n \circ \phi_{n-1} \circ \cdots \circ \phi_1$. Then the sequence of functions $\{f_n\}$ converges uniformly on any closed subset of R to a function f which has the properties we require: f is analytic on R, and defines a transformation which maps R one-one onto the open unit disc D. The verification of these assertions involves some hard analysis which is beyond the scope of this course. (A proof of the Riemann Mapping Theorem along the lines we have outlined is contained in Nevanlinna, R. and Paatero, V. *Introduction to Complex Analysis* Addison-Wesley, 1969.)

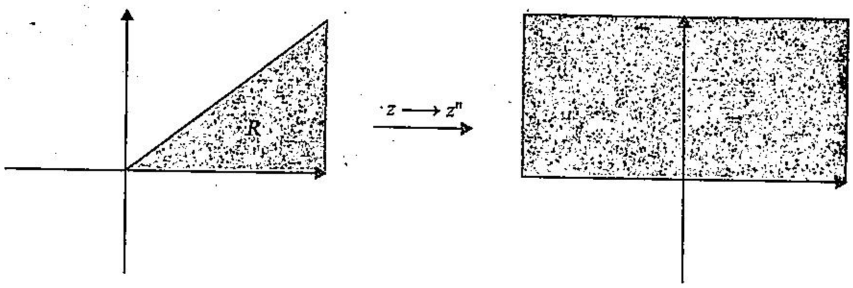
The practical difficulty about the Riemann Mapping Theorem is that though it tells us of the existence of an analytic equivalence of a simply-connected region with the open unit disc, it does not tell us how to find such an equivalence in a straightforward way. The best way of finding analytic equivalences of regions with the unit disc is to look at some of the familiar functions and see what they do. We give some examples below. More complicated regions may sometimes be dealt with by composition of several functions.

Examples

1. The bilinear transformation $z \longrightarrow \dfrac{1+iz}{1-iz}$ is an analytic equivalence of the upper half-plane $\{z:\operatorname{Im} z > 0\}$ with the unit disc $\{z:|z| < 1\}$. We discussed this transformation in Examples 1 and 4 of Section 13.3. The point of mentioning it again here is that it is often easier in practice to find an analytic equivalence of a given region with the upper half-plane than with the unit disc. An analytic equivalence with the unit disc can then be constructed by composition with this bilinear transformation.

2. The function $z \longrightarrow z^n$, where n is a positive integer, is an analytic equivalence of the region $\{z:0 < \operatorname{Arg} z < \pi/n\}$ with the upper half-plane. This region is a sector, of angle π/n. See Fig. 32.

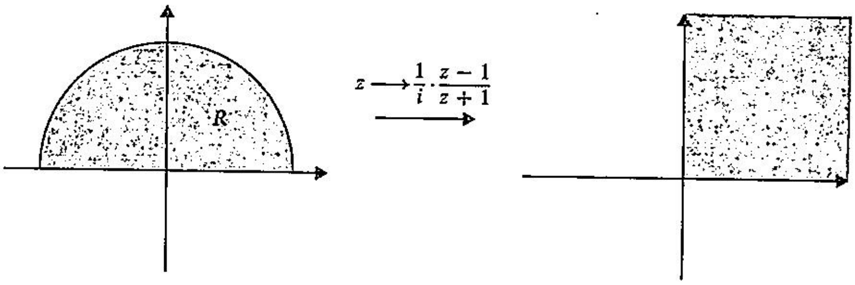
   In particular, the function $z \longrightarrow z^2$ is an analytic equivalence of the positive quadrant $\{z:\operatorname{Re} z > 0, \operatorname{Im} z > 0\}$ with the upper half-plane. The function $z \longrightarrow z^n$ fails to be conformal at 0 (which does *not* lie in the region in question). Evidently, whenever we have a region whose boundary has corners to map to the upper half-plane, the function required must fail to be conformal on the boundary of the region.



$R = \{z : 0 < \operatorname{Arg} z < \pi/n\}$

Fig. 32

3. The bilinear transformation $z \longrightarrow \frac{1}{i} \cdot \frac{z-1}{z+1}$ is an analytic equivalence of the half-disc $\{z : |z| < 1, \operatorname{Im} z > 0\}$ with the positive quadrant $\{z : \operatorname{Re} z > 0, \operatorname{Im} z > 0\}$.



$R = \{z : |z| < 1, \operatorname{Im} z > 0\}$

Fig. 33

It is the inverse of the bilinear transformation given in Example 1, so we know that it maps the whole disc $|z| < 1$ onto the whole upper half-plane. It maps the unit circle to the real axis. The images of the points 1, 0 and -1 under the bilinear transformation are the points 0, i, and ∞, respectively. The image of the real axis is thus the imaginary axis, and so the image of the half-disc is certainly one of the quadrants bounded by the real and imaginary axes. The image of the point $\frac{1}{2}i$, which lies in the half-disc, is $\frac{1}{5}(4 + 3i)$, which lies in the positive quadrant $\{z : \operatorname{Re} z > 0, \operatorname{Im} z > 0\}$. So the image of the half-disc is the positive quadrant.

In Example 2 we saw that $z \longrightarrow z^2$ is an analytic equivalence of the positive quadrant with the upper half-plane. The composition

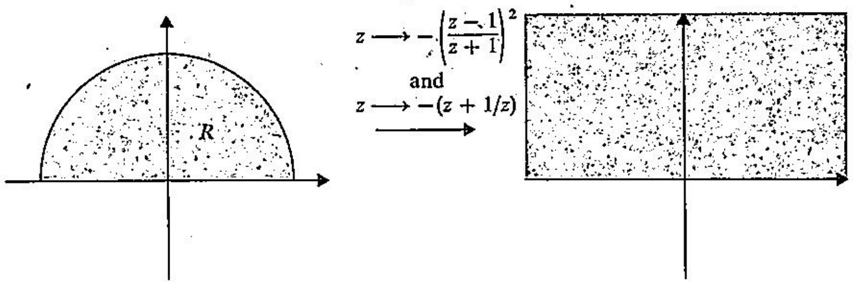
$$(z \longrightarrow z^2) \circ \left(z \longrightarrow \frac{1}{i} \cdot \frac{z-1}{z+1}\right) = z \longrightarrow -\left(\frac{z-1}{z+1}\right)^2$$

is therefore an analytic equivalence of the half-disc $\{z : |z| < 1, \operatorname{Im} z > 0\}$ with the upper half-plane. Notice that this time our analytic equivalence has a singularity on the boundary of the region.

We can obtain any other analytic equivalence of our region with the upper half-plane by composing the given one with an analytic equivalence of the upper half-plane with itself. For example, the bilinear transformation $z \longrightarrow 2\left(\frac{z-1}{z+1}\right)$ is an analytic equivalence of the upper half-plane with itself. (See Problem 4 of Section 13.4.) Thus the function

$$z \longrightarrow 2\left(-\left(\frac{z-1}{z+1}\right)^2 - 1\right) \Big/ \left(-\left(\frac{z-1}{z+1}\right)^2 + 1\right)$$

is an analytic equivalence of the half disc with the upper half-plane. This simplifies to $z \longrightarrow -(z + 1/z)$. See Fig. 34 on page 64.

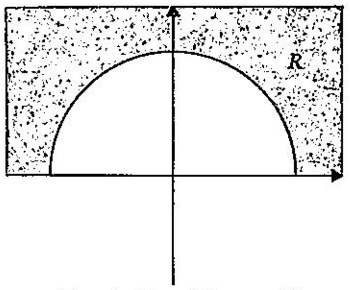

R
$z \longrightarrow -\left(\frac{z-1}{z+1}\right)^2$
and
$z \longrightarrow -(z + 1/z)$

$R = \{z : |z| < 1, \operatorname{Im} z > 0\}$

Fig. 34

4. We shall next construct an analytic equivalence of the region $\{z:|z| > 1, \operatorname{Im} z > 0\}$ with the upper half-plane. The region in question is that part of the upper half-plane lying outside the unit disc (Fig. 35).

   If we can first map it onto the half-disc $\{z:|z| < 1, \operatorname{Im} z > 0\}$, which looks pretty feasible, then we can use the function of Example 3 to map it onto the upper half-plane. The inversion $z \longrightarrow 1/z$ is an analytic equivalence of the outside of the unit circle with the inside of the unit circle, but unfortunately maps our region onto the lower half-disc. But this is easy to rectify: the function $z \longrightarrow -1/z$ is an analytic equivalence of our region with the upper half-disc. (One's immediate reaction is to use conjugation, that is to say, to think of the function $z \longrightarrow 1/\bar{z}$. But this function is not analytic, of course.) So the function obtained from the composition $(z \longrightarrow -1/z) \circ (z \longrightarrow -(z + 1/z))$ is an analytic equivalence of our region with the upper half-plane. This composite function is $z \longrightarrow z + 1/z$.

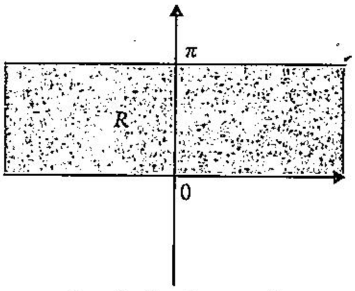


$R = \{z:|z| > 1, \operatorname{Im} z > 0\}$

Fig. 35

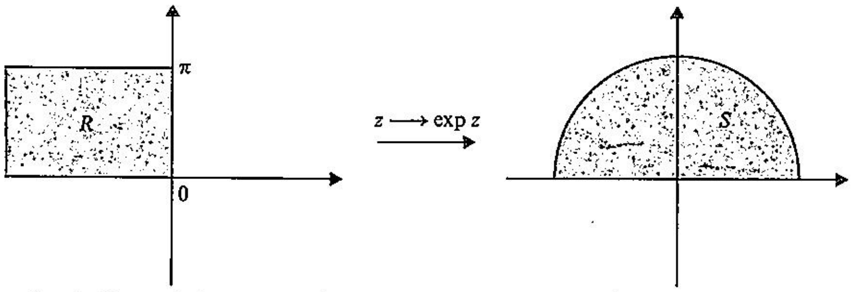
5. The exponential function may be used to give analytic equivalences of several interesting regions with the upper half-plane. For example, it is an analytic equivalence of the strip $\{z:0 < \operatorname{Im} z < \pi\}$ (Fig. 36) with the upper half-plane.

   It maps the half-strip $\{z:\operatorname{Re} z < 0, 0 < \operatorname{Im} z < \pi\}$ onto the half-disc $\{z:|z| < 1, \operatorname{Im} z > 0\}$.



$R = \{z:0 < \operatorname{Im} z < \pi\}$

Fig. 36

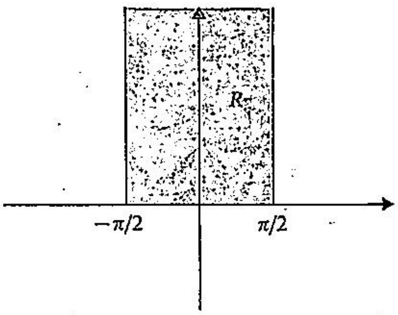


$R = \{z:\operatorname{Re} z < 0, 0 < \operatorname{Im} z < \pi\}$ $S = \{z:|z| < 1, \operatorname{Im} z > 0\}$

Fig. 37

So if we form the composition $(z \longrightarrow -(z + 1/z)) \circ \exp$, we obtain an analytic equivalence of the half-strip with the upper half-plane, namely $z \longrightarrow -(e^z + e^{-z})$. The insertion of a factor $\frac{1}{2}$ will do no harm, since $z \longrightarrow \frac{1}{2}z$ is an analytic equivalence of the upper half-plane with itself. Thus $z \longrightarrow -\cosh z$ is an analytic equivalence of the half-strip $\{z:\operatorname{Re} z < 0, 0 < \operatorname{Im} z < \pi\}$ with the upper half-plane.

In a similar vein, consider the vertical half-strip $\{z: -\pi/2 < \operatorname{Re} z < \pi/2, \operatorname{Im} z > 0\}$ (Fig. 38).



$R = \{z: -\pi/2 < \operatorname{Re} z < \pi/2, \operatorname{Im} z > 0\}$

Fig. 38

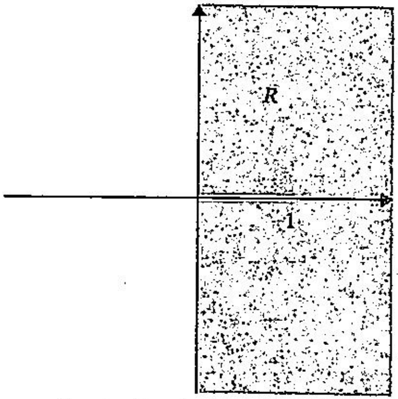
The function $z \longrightarrow iz + i\frac{\pi}{2}$ maps this vertical half-strip to our first, horizontal half-strip; so if we follow that function with $z \longrightarrow -\cosh z$ we obtain an analytic equivalence of the vertical half-strip with the upper half-plane. But

$$-\cosh\left(iz + i\frac{\pi}{2}\right) = -\cos\left(-z - \frac{\pi}{2}\right) = -\cos\left(z + \frac{\pi}{2}\right)$$
$$= \sin z.$$

So we have shown that the function $z \longrightarrow \sin z$ is an analytic equivalence of the region $\{z: -\frac{\pi}{2} < \operatorname{Re} z < \frac{\pi}{2}, \operatorname{Im} z > 0\}$ with the upper half-plane.

6. Here is a slightly more odd example. We consider the region $\{z: \operatorname{Re} z > 0,$ z is not a real number less than or equal to $1\}$ shown in Fig. 39. This region is the right-hand half-plane with a slit cut in it along the real axis from 0 to 1. We claim that the function $z \longrightarrow i\sqrt{z^2 - 1}$, where $\sqrt{\ }$ indicates the principal branch of the square root, is an analytic equivalence of this region with the upper half-plane.

   The function is the composition of several simpler functions. First, the function $z \longrightarrow z^2$ maps our region onto the plane slit along the real axis from the point 1 to the left. Then, the function $z \longrightarrow z - 1$ translates the slit plane so that the slit now lies along the non-positive real axis. The function $z \longrightarrow \sqrt{z}$ maps this region onto the right-hand half-plane; finally, the function $z \longrightarrow iz$ rotates the right-hand half-plane onto the upper half-plane.



$R = \{z: \operatorname{Re} z > 0, z$ not a real number less than or equal to $1\}$

Fig. 39

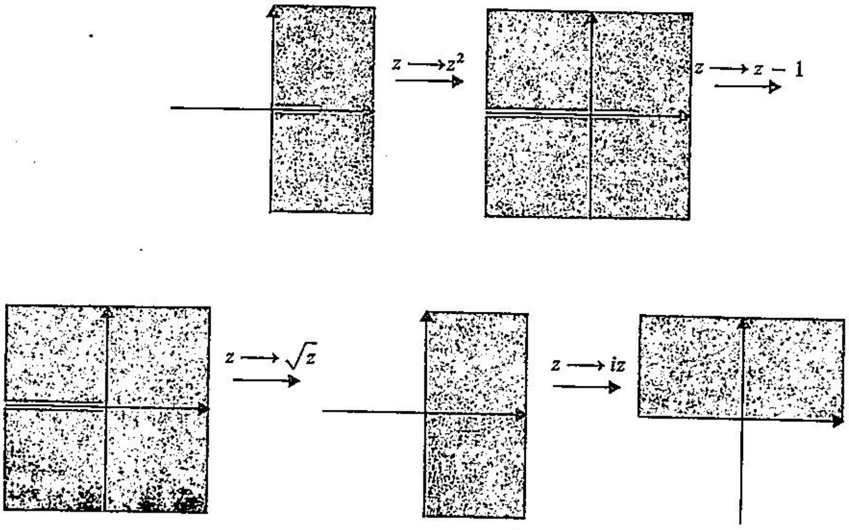
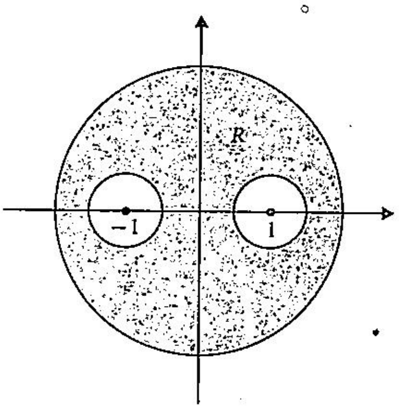


Fig. 40

Finally, what happens if we consider regions which are not simply-connected? The annulus is an example of such a region (it has one hole); the region $\{z: |z| < 2, |z - 1| > \frac{1}{2}, |z + 1| > \frac{1}{2}\}$, which has two holes, is another example of a region which is not simply-connected (Fig. 41).

It so happens that two regions related by a one-one function of the kind we have been discussing must have the same connectivity—that is, the same number of holes. (This result may be proved using topological methods. See *Unit M202 27, Connectedness.*) So no annulus can be analytically equivalent to the region which appears in Fig. 41. What about two regions with the same connectivity? It turns out that two annuli are analytically equivalent only if the ratio of external to internal radius is the same for each. So the situation for regions which are not simply-connected is complicated.



$R = \{z: |z| < 2, |z - 1| > \frac{1}{2}, |z + 1| > \frac{1}{2}\}$

Fig. 41

Summary

We have examined some questions about the existence of analytic equivalences between pairs of regions—that is, the existence of one-one analytic functions mapping one region onto the other. We have shown that the only region analytically equivalent to the whole complex plane is the complex plane itself. On the other hand, any pair of simply-connected regions both smaller than the complex plane are analytically equivalent. This is the content of the Riemann Mapping Theorem.

We have discussed some particular examples of analytic equivalences in detail.

Self-Assessment Questions

1. State, for each of the following regions R_1 and R_2, whether there is an analytic equivalence of R_1 with R_2. (You need not find such an equivalence.)
   (i) $R_1 = \mathbf{C}$; $R_2 = \mathbf{C} - \{0\}$.
   (ii) $R_1 = \{z:|z| < 1\}$; $R_2 = \{z:|z| > 1\}$.
   (iii) $R_1 = \{z:|z| < 1\}$; $R_2 = \{z:|z| < 1, z \neq 0\}$.
   (iv) R_1 = interior of the triangle whose vertices are 0, 1 and i;
   R_2 = interior of the ellipse $|z - 1| + |z + 1| = 4$.
2. We showed that the region $\{z:0 < \operatorname{Arg} z < \pi/n\}$ is analytically equivalent to the upper half-plane. A transformation defined by a one-one analytic function is necessarily conformal. Is there no contradiction between these two assertions?
3. Find explicitly an analytic equivalence of the half disc $\{z:|z| < 1, \operatorname{Im} z > 0\}$ with the strip $\{z:0 < \operatorname{Im} z < \pi\}$.

Solutions

1. (i) No. (ii) No (R_2 not simply-connected).
   (iii) No (R_2 not simply-connected). (iv) Yes.
2. The transformation need be conformal only on the interior of the region, not on its boundary.
3. Take a branch of the logarithm, L, which contains the upper half-plane in its domain; let $f(z) = L\left(\left(\frac{1+z}{1-z}\right)^2\right)$.

13.8 PROBLEMS

1. In Example 7 of Section 13.3 we showed that the bilinear transformations which map the open unit disc onto itself are those of the form $z \longrightarrow \dfrac{az+b}{\bar{b}z+\bar{a}}$ with $|b| < |a|$. In this question you are asked to show that these are the only analytic equivalences of the open unit disc with itself.

   (a) Suppose that f is an analytic equivalence of the open unit disc with itself such that $f(0) = 0$. Thus f is analytic, one-one, and $|f(z)| < 1$ when $|z| < 1$. Let $\phi(z) = f(z)/z$, $0 < |z| < 1$. Show that ϕ has a removable singularity at 0. Suppose the singularity removed. Show that if $|z| = r$, where $0 < r < 1$, then $|\phi(z)| < 1/r$, and deduce that $|\phi(z)| < 1/r$ for all z such that $|z| < r$. Conclude that $|\phi(z)| \leqslant 1$ for all z in the open unit disc.

   Apply a similar argument to the inverse of f to conclude that $|\phi(z)| \geqslant 1$ for all z in the open unit disc. Hence show that $f(z) = az$ where $|a| = 1$.

   (b) Show that any analytic equivalence of the open unit disc with itself is a bilinear transformation (of the type specified at the beginning of the question).

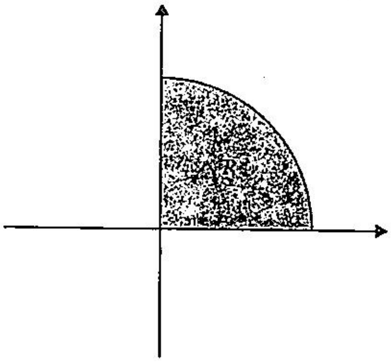
2. Let R be a simply-connected region in **C**, other than **C** itself, and suppose that f and g are analytic equivalences of R with the open unit disc. Suppose further that f and g map the same point z_0 to the centre of the unit disc, and that $f'(z_0)$ and $g'(z_0)$ are both real and positive. Show that $f = g$. (Hint: Use Problem 1).

   (This is a form of uniqueness theorem for the analytic equivalence defined in the Riemann Mapping Theorem—it says that the analytic equivalence of R with the unit disc is completely determined if one specifies which point is to be mapped to the origin, and that the analytic equivalence should have positive derivative at that point.)

3. Suppose that the function f maps the closed annulus $\{z: 1 \leqslant |z| \leqslant 2\}$ to the closed annulus $\{z: 1 \leqslant |z| \leqslant 4\}$, in such a way that f is continuous on $\{z: 1 \leqslant |z| \leqslant 2\}$, maps the circle $|z| = 1$ to itself, and the circle $|z| = 2$ to the circle $|z| = 4$, and is analytic on the open annulus $\{z: 1 < |z| < 2\}$. By applying the Maximum and Minimum Principles to the function $z \longrightarrow f(z)/z^2$, show that $|f(z)/z^2|$ is constant, and deduce that $f(z) = az^2$, where $|a| = 1$.

   Conclude that there is no analytic equivalence of the open annulus $\{z: 1 < |z| < 2\}$ with the open annulus $\{z: 1 < |z| < 4\}$ which can be extended continuously to the boundaries of the annuli in such a way that boundary is mapped to boundary.

4. Explain why the function $z \longrightarrow z^4$ is not an analytic equivalence of the quarter-disc $\{z: \text{Re}\, z > 0,\ \text{Im}\, z > 0 \text{ and } |z| < 1\}$ (Fig. 42) with the unit disc $\{z: |z| < 1\}$. Find an analytic equivalence of the quarter-disc with the unit disc.



$R = \{z: \text{Re}\, z > 0, \text{Im}\, z > 0$ and $|z| < 1\}$

Fig. 42

Solutions

1. (a) Since f has a (necessarily simple) zero at 0, we may write $f(z) = zg(z)$ for z in some neighbourhood of 0, where g is analytic (and $g(0) \neq 0$). Thus $\lim_{z\to 0} \frac{f(z)}{z} = g(0)$ exists, and ϕ, where $\phi(z) = f(z)/z$, has a removable singularity at 0.

Now if $|z| = r$, then

$$|\phi(z)| = \frac{|f(z)|}{r} < \frac{1}{r}, \quad \text{since } |f(z)| < 1.$$

Since ϕ is analytic on the disc $|z| < 1$ (assuming the singularity at 0 is removed), we conclude from the Maximum Principle that

$$|\phi(z)| < \frac{1}{r} \quad \text{for all } z \text{ such that } |z| \leqslant r.$$

So given any point z in the open unit disc, we know that $|\phi(z)| < 1/r$ for all r such that $|z| \leqslant r < 1$. It follows that $|\phi(z)| \leqslant 1$. (For suppose that $|\phi(z)| > 1$ for some z such that $|z| < 1$. Then $1/|\phi(z)| < 1$. Let r be any number less than 1 but greater than $\max\{|z|, 1/|\phi(z)|\}$. Then $r > 1/|\phi(z)|$, so $|\phi(z)| > 1/r$, even though $|z| < r < 1$, which is a contradiction.)

Now f^{-1} is also an analytic equivalence of the open unit disc with itself, so if we set $\psi(z) = f^{-1}(z)/z$, then $|\psi(z)| \leqslant 1$ for all z such that $|z| < 1$. But

$$\psi(f(z)) = \frac{z}{f(z)} = \frac{1}{\phi(z)},$$

so $|1/\phi(z)| \leqslant 1$, or $|\phi(z)| \geqslant 1$, for all z such that $|z| < 1$. We conclude that $|\phi(z)| = 1$ for all z in the open unit disc. But ϕ is analytic, and an analytic function whose modulus is constant must itself be constant (by Theorem 17 of *Unit 3*). Thus $\phi(z) = a$, where $|a| = 1$, that is, $f(z) = az$.

(b) Let f be an analytic equivalence of the open unit disc with itself. If $f(0) = 0$, then f is bilinear, by part (a). If not, let $f(0) = w_0$, where $|w_0| < 1$. Let $g(z) = \frac{z - w_0}{-\overline{w}_0 z + 1}$. Then g is a bilinear transformation which maps the unit disc to itself, and $g(f(0)) = 0$. Thus $g \circ f$ is an analytic equivalence of the open unit disc with itself which leaves the origin fixed. Thus $(g \circ f)(z) = az$, where $|a| = 1$, and so $f(z) = g^{-1}(az)$, which means that f is a bilinear transformation of the appropriate type.

2. The function $f \circ g^{-1}$ is an analytic equivalence of the unit disc with itself, and

$$(f \circ g^{-1})(0) = f(z_0) = 0.$$

Thus by part (a) of Problem 1, $(f \circ g^{-1})(z) = az$, where $|a| = 1$, and so $f(z) = a \cdot g(z)$. Thus $f'(z_0) = a \cdot g'(z_0)$; so if $f'(z_0)$ and $g'(z_0)$ are both real and positive, then a must be 1. So $f = g$.

3. The function $z \longrightarrow |f(z)/z^2|$ has the value 1 on the boundary of the annulus $\{z : 1 \leqslant |z| \leqslant 2\}$, since if $|z| = 1$ then $|f(z)| = 1$, whereas if $|z| = 2$ then $|f(z)| = 4$. The function $z \longrightarrow f(z)/z^2$ is analytic on the open annulus and continuous on its closure, and so by the Maximum Principle, $|f(z)/z^2| \leqslant 1$ for all z in the open annulus. On the other hand, $z \longrightarrow f(z)/z^2$ is also non-zero on the closed annulus, so by the Minimum Principle, $|f(z)/z^2| \geqslant 1$ for all z in the open annulus. Thus $|f(z)/z^2| = 1$ for all z in the open annulus. An analytic function with constant modulus is constant (by Theorem 17 of *Unit 3*), so $f(z) = az^2$, where $|a| = 1$.

Since the function $z \longrightarrow az^2$ is not an analytic equivalence of the two open annuli (because it is not one-one), there is no analytic equivalence of $\{z : 1 < |z| < 2\}$ with $\{z : 1 < |z| < 4\}$ which extends continuously to their boundaries.

4. The function $z \longrightarrow z^4$ is an analytic equivalence of the quarter disc with the unit disc *with the positive real axis removed.* In other words, it fails to be an analytic equivalence of the quarter-disc with the unit disc because the image of the quarter-disc is not the whole unit disc. To find an analytic equivalence, we proceed by stages. The function $f:z \longrightarrow z^2$ is an analytic equivalence of the quarter-disc with the half-disc $\{z:\text{Im}\, z > 0, |z| < 1\}$. The function $g:z \longrightarrow -(z + 1/z)$ is an analytic equivalence of the half-disc with the upper half-plane. (See Example 3, page 63.) The bilinear transformation $h:z \longrightarrow \dfrac{1+iz}{1-iz}$ is an analytic equivalence of the upper half-plane with the unit disc. (See Example 1, page 62.) Thus the composition $h \circ g \circ f$ will do the trick. This composition is

$$z \longrightarrow \frac{1 - i\left(z^2 + \dfrac{1}{z^2}\right)}{1 + i\left(z^2 + \dfrac{1}{z^2}\right)},$$

which simplifies (slightly) to

$$z \longrightarrow \frac{-iz^4 + z^2 - i}{iz^4 + z^2 + i}.$$

13.9 NON-EUCLIDEAN GEOMETRY

Until the first quarter of the nineteenth century, geometry came in only one variety, that first formalized by Euclid in about 300 B.C.

Euclid's influence has been staggering. Geometry, taught in what one imagines was much the same spirit as he taught it, has been taught in schools up to the present day. This phenomenon may perhaps be accounted for by the nature of Euclidean argument. The axiomatic system, which was the most significant contribution of the Greeks to mathematics, reached its mature form in Euclid's "Elements". Each proposition in the "Elements" was deduced from previously established propositions, or from certain basic rules, which were regarded as so self-evident as to require no justification. These basic rules we call *axioms*, and a theory built up in the way we have just described is called an *axiomatic system*. (The axioms must be supplemented by definitions of the technical terms used in the theory, of course.) This style of argument is familiar to anyone who studies modern mathematics. Although we have not been particularly careful to emphasize the fact, this complex analysis course follows the same pattern: each result is the end of a chain of reasoning that begins with the basic properties of the complex numbers, and the definitions of continuous function, analytic function, and so on.

Here are Euclid's five geometrical axioms (there are some others, which are essentially algebraic in character, which do not concern us):

1. A straight line can be drawn between any two points.
2. A finite straight line can be extended to any length in a straight line in either direction.
3. A circle may be described with any centre and any radius.
4. All right angles are equal.
5. If a straight line transverse to two given straight lines makes interior angles on the same side together less than two right angles, the two given straight lines will meet on the side on which the angles are together less than two right angles.

The fifth axiom is equivalent to the following:

> Through a given point not lying on a given straight line there can be drawn exactly one straight line parallel to the given line.

Parallel lines are lines which, being produced indefinitely in both directions, do not meet one another in either direction. Euclid's fifth axiom is often referred to as the *parallel axiom*.

From the beginning, the parallel axiom was regarded with suspicion. The validity of an axiom should be self-evident. Many people, when examining the Euclidean system critically, failed to find the parallel axiom self-evident. The natural thing to do, when faced with a statement which is not self-evident, is to try to prove it; much effort was spent, particularly in the eighteenth century, in trying to prove the parallel axiom from the remaining ones. This effort was in vain. Mathematicians eventually discovered that the parallel axiom is independent of the other Euclidean axioms (and so cannot be proved from them) and realized that there is no reason, at least on the abstract level, why one should not investigate those *abstract geometries* (or axiomatic systems) which arise if the parallel axiom is replaced by something else.

There are two obvious alternatives to the Euclidean parallel axiom:

> through a given point not lying on a given straight line no straight line can be drawn parallel to the given line;

or

> through a given point not lying on a given straight line more than one straight line can be drawn parallel to the given line.

The first alternative turns out to be inconsistent with the second Euclidean axiom; in such a geometry, lines cannot be extended to arbitrary length. But the second alternative gives a self-consistent geometry, which differs considerably from Euclidean geometry. The development of this geometry was begun independently and at about the same time by three men: the German Karl Friedrich Gauss (1777–1855), the Hungarian Janos Bolyai (1802–1860), and the Russian Nicolai Ivanovitch Lobachevsky (1793–1856). It was Lobachevsky who was first to publish his results (in 1829–30), and so the new geometry became known as **Lobachevskian geometry**. It was the first example of Non-Euclidean geometry to be discovered.

Once the principle had been admitted, that it is permissible to investigate geometries with axioms different from the Euclidean ones, then more Non-Euclidean geometries, with other changes in the axioms, were discovered. For example, Bernhard Riemann showed in 1854 that there is a Non-Euclidean geometry in which there are no parallels, if the second of Euclid's axioms is also adjusted.

The work of Lobachevsky, Bolyai and Gauss was somewhat abstract. It contained no clue as to how Lobachevskian geometry could be realized or pictured. Not long after its discovery, however, various models of Lobachevskian geometry were proposed. In the remainder of this section we shall describe a model of Lobachevskian geometry discovered by the Frenchman Henri Poincaré (1854–1912) which uses some of the ideas about transformations which we have developed in this unit.

A *model of an abstract geometry* consists of a set whose points (or elements) are to be identified with the "points" of the geometry, a collection of lines in the set which are to be identified with the "straight lines" of the geometry, and a notion of angle and length in the model, all chosen in such a way as to satisfy the axioms of the geometry. As a rather trivial example, we can provide a model of *Euclidean* geometry in which the "points" of the geometry are* the points of the complex plane; the "straight lines" of the geometry are the paths of arcs of the form $t \longrightarrow \alpha t + \beta, \alpha \neq 0, t$ in some closed interval $[a, b]$; the "angle" between two "straight lines" is the angle between the arcs $t \longrightarrow \alpha_1 t + \beta$ and $t \longrightarrow \alpha_2 t + \beta$ at β, which is $\operatorname{Arg} \alpha_2 - \operatorname{Arg} \alpha_1$, up to multiples of 2π; the "length" of the "straight line" joining z_1 and z_2 is $|z_1 - z_2|$. It should be clear that the axioms of Euclidean geometry are satisfied with these definitions. There is just one parallel to the straight line defined by $t \longrightarrow \alpha t + \beta$ which passes through the point γ, namely the straight line defined by $t \longrightarrow \alpha t + \gamma$.

In this case it takes an effort of will to distinguish between the geometry and its model. We can, however, give a more serious example.

It is fairly easy to give a model of Riemann's Non-Euclidean geometry—the one in which there are no parallels. We take the "points" of the geometry to be the points of the unit sphere; the "straight lines" to be arcs of great circles of the sphere (a great circle is the intersection of the sphere with a plane through its centre); the "angle" between two "straight lines" is the angle between the two corresponding arcs, that is between their tangents at the point of intersection; the "length" of a line segment is the angle the corresponding arc subtends at the centre of the sphere. Since any two great circles intersect (at the opposite ends of the diameter along which the planes that define them intersect), there are no parallels in this model.

We could equally well have taken the extended complex plane as a model of Riemann's (Non-Euclidean) geometry: but then the definition of "length" becomes a little tricky. But the exercise of translating the sphere model into a model in the extended complex plane is valuable because it suggests a way of modelling Lobachevskian geometry in the complex plane. For example, we know that the bilinear transformations of the form $z \longrightarrow az + b$ with $|a| = 1$ are important in Euclidean geometry since they are the isometries of the plane onto itself (page 23). The bilinear transformations of the form $z \longrightarrow \dfrac{az + b}{-\bar{b}z + \bar{a}}$ play the same role in spherical geometry, since they represent the rotations of the sphere (page 52). We shall construct our model of Lobachevskian geometry out of yet another family of bilinear transformations, and by analogy with the extended complex plane model of Riemann's geometry.

---

* Strictly, we should say: "are identified with". The same remark applies in the rest of the description of this model and subsequent models.

Extended Complex Plane Model of Riemann's Geometry

We begin by examining this model of Riemann's geometry. (We shall use SMALL CAPITALS for the items of the model we are trying to define.)

The POINTS in our new model of Riemann's geometry are just the points of the complex plane, together with the point at infinity. The LINES of the model are circles and lines in the plane of a certain type, namely those whose stereographic images are great circles on the sphere. The lines we need are those that pass through the origin. It would be possible, but not particularly profitable, to give a characterization of the circles in the plane corresponding to great circles on the sphere in terms of plane geometry. Since stereographic projection preserves angles, and angles in the sphere model are the usual angles, ANGLES in the new model coincide with the usual angles. The difficulty is to define LENGTH in the new model, but again the sphere model will guide us.

Let γ be an arc whose path is a LINE of the model. The length of the path of γ from t_1 to t_2 is $\int_{t_1}^{t_2} |\gamma'(t)|\, dt$, or $\int_{\gamma} |dz|$. But this is not the LENGTH of γ. (Observe the strategic small capitals.) Perhaps the best way to understand what is going on is to think about the stereographic image of the path of γ on the sphere. The point is that stereographic projection distorts distances. What we want to do is to find how to express, in terms of γ, the length of the *stereographic image* of its path. We can use our knowledge of the bilinear transformations which correspond to rotations of the sphere to work out how to do it.

We may assume that the LENGTH of the path of γ is $\int_{\gamma} \phi(z)|dz|$, where ϕ is some continuous positive real-valued function on the extended complex plane. The inclusion of the factor ϕ compensates for the differing ratios between distances on the sphere and on the plane. How do we calculate ϕ? One thing is certain. If f is a bilinear transformation of the type that corresponds to a rotation of the sphere, then the LENGTH of the path of $f(\gamma)$ from t_1 to t_2 must be the same as the LENGTH of the path of γ from t_1 to t_2, since rotations preserve "lengths". To determine ϕ we need consider only a particular case for γ. Let $\gamma(t) = t$; then the path of γ is a LINE. Let ζ be any point of the plane. The bilinear transformation $f(z) = \dfrac{z + \zeta}{-\bar{\zeta}z + 1}$ corresponds to a rotation, and maps 0 to ζ. The image of γ under this bilinear transformation is $f(\gamma)$, where $f(\gamma(t)) = \dfrac{t + \zeta}{-\bar{\zeta}t + 1}$. The LENGTH of $f(\gamma)$ from 0 to t is

$$\int_{f(\gamma)} \phi(z)|dz| = \int_0^t \phi\left(\frac{t + \zeta}{-\bar{\zeta}t + 1}\right) \cdot \frac{1 + |\zeta|^2}{(-\bar{\zeta}t + 1)^2} dt.$$

The LENGTH of γ from 0 to t is

$$\int_{\gamma} \phi(z)|dz| = \int_0^t \phi(t)\, dt.$$

These LENGTHS are equal, so

$$\int_0^t \phi\left(\frac{t + \zeta}{-\bar{\zeta}t + 1}\right) \cdot \frac{1 + |\zeta|^2}{(-\bar{\zeta}t + 1)^2} dt = \int_0^t \phi(t)\, dt.$$

Since these integrals are equal for all t, and the integrands are continuous, the integrands must be equal; in particular, they must have the same value when $t = 0$. Thus $\phi(\zeta) \cdot (1 + |\zeta|^2) = \phi(0)$, and so

$$\phi(\zeta) = \frac{\phi(0)}{1 + |\zeta|^2}.$$

Thus we are able to determine ϕ precisely once we know its value at 0. In a sense it is not very important to work out this final detail, since it amounts only to defining a unit of LENGTH, and in any case $\phi(0)$ depends on the radius of the sphere. But it is very simple to do, and the method illustrates how this measurement of LENGTH works in practice, so we might as well do it. We shall calculate the total LENGTH of the curve $\gamma(t) = t$. This curve (the real axis) corresponds to a great circle of the sphere, so its LENGTH should be $2\pi r$, where r is the radius of the sphere. Well,

$$\int_\gamma \phi(z)|dz| = \int_{-\infty}^{\infty} \frac{\phi(0)}{1 + t^2} dt = \phi(0) \cdot 2 \arctan t \Big|_{-\infty}^{\infty}$$

$$= \phi(0) \cdot \pi.$$

Thus $\phi(0) = 2r$. We see again that $\frac{1}{2}$ is a natural choice of radius for the sphere: not only does it make the equator correspond to the unit circle in the plane (see Problem 3(b) of Section 13.6), it also makes $\phi(0) = 1$.

Notice that $\lim_{z \to \infty} \phi(z) = 0$. This reflects the fact that points in the plane far from the origin correspond to points on the sphere close to the north pole, and which are close together.

One final remark about all this, before we turn to Lobachevskian geometry. We have talked only of the LENGTHS of LINES in the model. But any smooth arc in the plane has a stereographic image on the sphere, and so has a LENGTH, and the LENGTH of *any* smooth arc γ is given by the same formula, $\int_\gamma \phi(z)|dz|$.

Open Unit Disc Model of Lobachevskian Geometry

The POINTS of our model of Lobachevskian geometry are the points of the open unit disc in the complex plane, that is, the points $\{z : |z| < 1\}$. ANGLE in our model again coincides with the usual notion of angle. We could say that our model is a conformal one. We leave aside for a moment the description of the LINES of the model, and consider first the definition of LENGTH.

We shall construct a measure of LENGTH in our model of Lobachevskian geometry in much the same way that we constructed LENGTH in Riemann's geometry, but using a different family of basic transformations. We shall assume that LENGTH in our model of Lobachevskian geometry is preserved by all analytic equivalences of the unit disc with itself. An analytic equivalence of the unit disc with itself is necessarily a bilinear transformation of the form

$$z \longrightarrow \frac{az + b}{\bar{b}z + \bar{a}}, \text{ where } |b| < |a|. \text{ (See Problem 1 of Section 13.8.)}$$

We assume that the LENGTH of any smooth arc γ is given by $\int_\gamma \psi(z)|dz|$, where ψ is some positive real-valued function whose domain is the unit disc. If f is an analytic equivalence of the unit disc with itself, then $f(\gamma)$ is to have the same LENGTH as γ. Let $\gamma(t) = t$, and let $f(z) = \dfrac{z + \zeta}{\bar{\zeta}z + 1}$, where $|\zeta| < 1$. Then f is an analytic equivalence of the unit disc with itself, and maps the point 0 to the point ζ. The LENGTH of $f(\gamma)$ from 0 to t is

$$\int_{f(\gamma)} \psi(z)|dz| = \int_0^t \psi\left(\frac{t + \zeta}{\bar{\zeta}t + 1}\right) \cdot \frac{1 - |\zeta|^2}{(\bar{\zeta}t + 1)^2} dt.$$

The LENGTH of γ from 0 to t is

$$\int_\gamma \psi(z)|dz| = \int_0^t \psi(t)\,dt.$$

It follows as before, by setting $t = 0$, that

$$\psi(\zeta)\cdot(1 - |\zeta|^2) = \psi(0).$$

Thus $\psi(\zeta)$ is defined once the value of $\psi(0)$ is decided. Again, the value of $\psi(0)$ is arbitrary, and this time we have no external considerations to help us choose it. So we take $\psi(0)$ to be 1, because it is the most obvious choice. Thus

$$\psi(z) = \frac{1}{1 - |z|^2}, \quad \text{where } |z| < 1.$$

Notice the crucial minus sign. One of its effects is that ψ is very large for points near the boundary of the unit disc (for such points $|z|$ is close to 1). Thus points close together (in ordinary measure) near the boundary of the unit disc are far apart in terms of the Lobachevskian measure of distance. It is as though some unnoticed physical effect makes the traveller go ever more slowly, as he approaches the boundary.

Strictly speaking we should complete our discussion of the function ψ by showing that the LENGTH it defines really is invariant under all analytic equivalences of the unit disc with itself, and not just under the special ones we used in its derivation. We shall leave this as an exercise (Problem 2 of Section 16.10) and pass on to the determination of the LINES of our model.

The idea that we shall exploit in finding the LINES in our model of Lobachevskian geometry is that the LENGTH of a LINE should give the shortest DISTANCE between two POINTS. The shortest distance between two points of the plane is the length of the straight line joining them, and the shortest distance between two points on the sphere is the length of the shorter arc of the great circle through them. We begin our investigations of the corresponding curves in Lobachevskian geometry by showing that of all the smooth arcs joining the points 0 and r (where r is a positive real number less than 1) whose paths lie inside the unit disc, the one with the shortest LENGTH is the segment of the real axis from 0 to r.

Let us calculate the LENGTH of the segment of the real axis from 0 to r (Fig. 43). It is

$$\int_0^r \psi(t)\,dt = \int_0^r \frac{1}{1-t^2}dt = \frac{1}{2}\log\frac{1+t}{1-t}\bigg|_0^r$$
$$= \frac{1}{2}\log\frac{1+r}{1-r}.$$

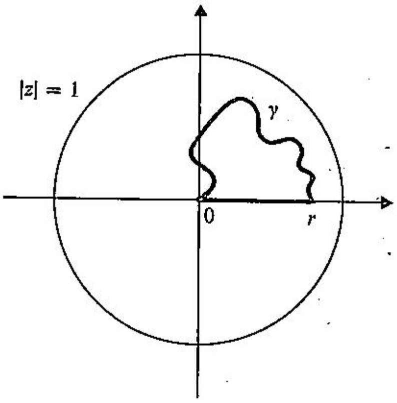


Fig. 43

Now let γ be a smooth arc such that $\gamma(0) = 0$, $\gamma(T) = r$, and $|\gamma(t)| < 1$ for all t in $[0, T]$. The LENGTH of γ is

$$\int_0^T \psi(\gamma(t))|\gamma'(t)|dt = \int_0^T \frac{|\gamma'(t)|}{1 - |\gamma(t)|^2}dt.$$

Let $\gamma(t) = u(t) + iv(t)$, where u and v are real-valued. Then

$$|\gamma(t)|^2 \geqslant [u(t)]^2 \quad \text{for all } t \text{ in } [0, T],$$

so

$$\frac{1}{1 - |\gamma(t)|^2} \geqslant \frac{1}{1 - [u(t)]^2} \quad \text{for all } t \text{ in } [0, T].$$

Moreover,

$$\gamma'(t) = u'(t) + iv'(t),$$

so

$$|\gamma'(t)| \geqslant |u'(t)| \geqslant u'(t).$$

Thus

$$\begin{aligned}\int_0^T \frac{|\gamma'(t)|}{1-|\gamma(t)|^2}\,dt &\geqslant \int_0^T \frac{u'(t)}{1-[u(t)]^2}\,dt\\ &= \int_0^r \frac{1}{1-U^2}\,dU, \quad \text{substituting } U = u(t),\\ &= \frac{1}{2}\log\frac{1+r}{1-r}.\end{aligned}$$

Thus the LENGTH of γ is at least as great as the LENGTH of the line segment $[0, r]$ of the real axis. But we have to show that the LENGTH of γ actually exceeds the LENGTH of the segment, unless γ is a proper reparametrization of the segment. Now the equality

$$\int_0^T \frac{|\gamma'(t)|}{1-|\gamma(t)|^2}\,dt = \frac{1}{2}\log\frac{1+r}{1-r}$$

can hold only if

$$\int_0^T \frac{|\gamma'(t)|}{1-|\gamma(t)|^2}\,dt = \int_0^T \frac{u'(t)}{1-[u(t)]^2}\,dt.$$

If either $|\gamma(t)|^2 > [u(t)]^2$ for some t, or $|\gamma'(t)| > u'(t)$ for some t, then since the integrands are continuous, it would be the case that

$$\int_0^T \frac{|\gamma'(t)|}{1-|\gamma(t)|^2}\,dt > \int_0^T \frac{u'(t)}{1-[u(t)]^2}\,dt.$$

Thus for equality to hold, we must have

$$|\gamma(t)|^2 = [u(t)]^2 \quad \text{and} \quad |\gamma'(t)| = u'(t) \quad \text{for all } t \text{ in } [0, T].$$

But then $\gamma(t) = u(t)$, and $u'(t) > 0$, for all t in $[0, T]$, in which case γ is a proper reparametrization of the segment.

We have shown that the LINE joining the points 0 and r, where r is real and positive, is the line segment $[0, r]$ of the real axis. To find the LINE joining another pair of points, we make use again of the analytic equivalence of the unit disc with itself. Since these analytic equivalences preserve LENGTH, they also map LINES to LINES. That is to say, if γ is the LINE joining z_1 to z_2 and f is an analytic equivalence of the unit disc with itself, then $f(\gamma)$ is the LINE joining $f(z_1)$ to $f(z_2)$. (If you do not find this obvious, try Problem 3 of Section 13.10.) Now given any pair of points z_1, z_2 in the unit disc, there is an analytic equivalence f of the unit disc with itself such that $f(z_1) = 0$ and $f(z_2)$ lies on the positive real axis. The LINE joining z_1 and z_2 is then the image of the line segment $[0, f(z_2)]$ of the real axis under f^{-1}. This will be a circle or straight line through z_1 and z_2, since f is a bilinear transformation. To specify it exactly, notice that the real axis is the unique member of the set of circles and lines through 0 and r, which intersects the unit circle at right angles. Bilinear transformations are of course conformal. We conclude that the LINE through z_1 and z_2 is the unique circle or straight line through z_1 and z_2 which intersects the unit circle at right angles (Fig. 44).

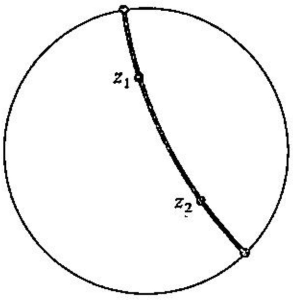


Fig. 44

The LENGTH of the LINE joining 0 and r, where r is a positive real number is, $\frac{1}{2}\log\frac{1+r}{1-r}$. Now $\frac{1+r}{1-r}$ is the cross-ratio of the points $1, 0, -1, r$; and 1 and -1 are points in which the LINE intersects the unit circle. By Problem 5 of Section 13.4 we know that cross-ratio is preserved by bilinear transformations. Thus the LENGTH of the LINE joining z_1 and z_2 is

$$\frac{1}{2}\log\frac{(\zeta_1 - z_1)(\zeta_2 - z_2)}{(z_1 - \zeta_2)(z_2 - \zeta_1)},$$

where ζ_1 and ζ_2 are the points in which the LINE meets the unit circle, as shown in Fig. 45 (the order is important).

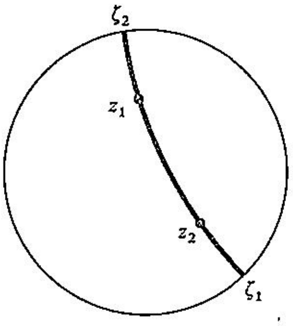


Fig. 45

Notice that as z_2 approaches ζ_1, so $\frac{1}{2}\log\frac{(\zeta_1 - z_1)(\zeta_2 - z_2)}{(z_1 - \zeta_2)(z_2 - \zeta_1)}$ becomes unbounded. The boundary of the unit disc is infinitely "distant" from each point inside it.

Now consider a LINE PQ and a point R not on it. (Fig. 46).

The pairs of LINES PQ and PR, and PQ and RQ, meet on the boundary of the unit disc, which is infinitely "distant" from R; in other words, they are parallel pairs of lines so far as the model is concerned. Moreover, any line through R lying in the angle marked in Fig. 46 also fails to meet PQ, and so is parallel to it. So the parallel axiom which holds in our geometry is indeed the Lobachevskian one.

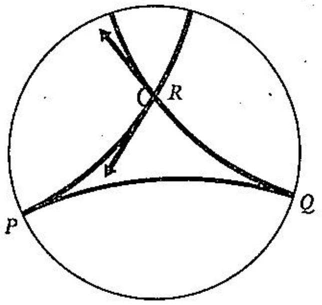


Fig. 46

Summary

Non-Euclidean geometry began when it was realized that the parallel axiom of Euclidean geometry was independent of the other axioms, and that there is no reason why one should not investigate geometries with different parallel axioms. It is possible to construct a model of Lobachevskian geometry in which the points are points of the unit disc in the complex plane, and the line joining two points is the arc of the circle through the points which is orthogonal to the unit circle. The model is conformal: angles are the same as Euclidean angles—but the distance between two points is given by the logarithm of a certain cross-ratio. There are strong similarities between the method of construction of this model, and the construction of a model of Riemann's geometry in the complex plane. In particular, the bilinear transformation of the unit disc onto itself play the role of isometries, or rotations of the sphere.

Self-Assessment Questions

1. Find the DISTANCE between the points $-\frac{1}{2}$ and $\frac{1}{2}$ in
   (i) Euclidean geometry;
   (ii) Riemann's geometry;
   (iii) Lobachevskian geometry.
2. Draw a rough diagram of the open unit disc, the Lobachevskian LINE L joining the points 0 and $\frac{1}{2}i$, and the point $\frac{1}{2}$. Indicate where the Lobachevskian parallels to L through $\frac{1}{2}$ lie.

Solutions

1. (i) 1.

   (ii) $\int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{1+t^2}\,dt = 2\arctan\frac{1}{2}$.

   (iii) $\frac{1}{2}\log\frac{(1+\frac{1}{2})(-1-\frac{1}{2})}{(-\frac{1}{2}+1)(\frac{1}{2}-1)} = \frac{1}{2}\log\frac{(\frac{3}{2})^2}{(\frac{1}{2})^2} = \log 3.$

2.

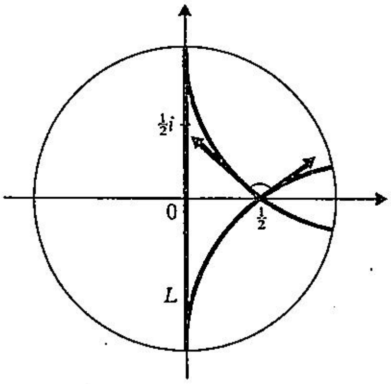


Fig. 47

The LINE L is the imaginary axis.

13.10 PROBLEMS

1. The following statements are true, and express what we hope are familiar facts, in Euclidean geometry. Are they true in Lobachevskian geometry?

   (i) The angles of a triangle sum to π.

   (ii) The corresponding angles formed by a transversal to two parallel lines are equal (Fig. 48).

   (iii) If A, B and C lie on a line, with B between A and C, then the length of AC is the sum of the length of AB and the length of BC.

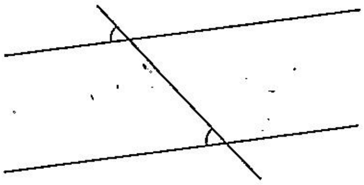


Fig. 48

2. Show that if $\psi(z) = \dfrac{1}{1 - |z|^2}$, where $|z| < 1$, $\gamma : [0, T] \longrightarrow D$ is any smooth arc in the open disc D, and f is any analytic equivalence of D with itself, then

$$\int_{f(\gamma)} \psi(z)|dz| = \int_{\gamma} \psi(z)|dz|.$$

In other words, show that LENGTH in Lobachevskian geometry is preserved by analytic equivalences of the unit disc with itself.

3. Show that if γ is the LINE joining z_1 and z_2, and f is an analytic equivalence of the unit disc with itself, then $f(\gamma)$ is the LINE joining $f(z_1)$ and $f(z_2)$.

4. Let us denote by $L(P, Q)$ the Lobachevskian LENGTH of the LINE joining the points P and Q.

   (a) Show that our formula for $L(O, P)$, when O is the origin and P a point of the real axis such that $OP = r$ (in the Euclidean measure of length), can be equivalently written $r = \tanh L(O, P)$. Deduce that for any point Q,

$$\tanh L(O, Q) = OQ,$$

   where OQ is the Euclidean distance from O to Q.

   (b) Prove the following "Pythagorean" theorem for the right-angled triangle shown in Fig. 49:

$$\cosh 2L(P, Q) = \cosh 2L(O, P) \cdot \cosh 2L(O, Q).$$

   Deduce that a similar formula is true for any right-angled triangle in Lobachevskian geometry. (To calculate the LENGTH of PQ, use an analytic equivalence of the unit disc with itself to bring the point P, say, to the origin.)

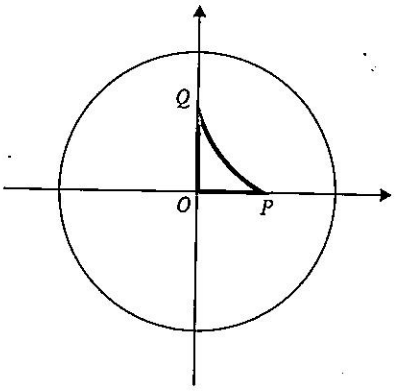
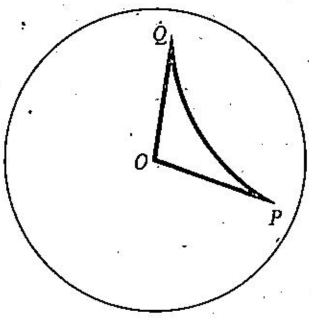


Fig. 49

Solutions



1. (i) False. Consider for example the triangle OPQ shown in Fig. 50. Its ANGLES evidently have sum less than π. In fact, any triangle may be moved, by means of an analytic equivalence of the unit disc with itself, to a triangle with one vertex at the origin, which therefore looks like the one in the figure. Since ANGLES are preserved (the mapping is conformal), we conclude that for every triangle the ANGLE sum is less than π.

   (ii) False. Since there are many parallels to a given LINE through a given point not on it, we could hardly expect this to be true.

   (iii) True. LENGTH is defined in terms of an integral, which is additive.

2. We may write $f(z) = \dfrac{az + b}{\bar{b}z + \bar{a}}$, with $|b| < |a|$. Then

$$\begin{aligned}(\psi \circ f)(z) &= \frac{1}{1 - |f(z)|^2}\\ &= \frac{(\bar{b}z + \bar{a})(b\bar{z} + a)}{(\bar{b}z + \bar{a})(b\bar{z} + a) - (az + b)(\bar{a}\bar{z} + \bar{b})}\\ &= \frac{(\bar{b}z + \bar{a})(b\bar{z} + a)}{(|a|^2 - |b|^2)(1 - |z|^2)};\end{aligned}$$

also

$$\begin{aligned}|f'(z)| &= \left|\frac{a(\bar{b}z + \bar{a}) - \bar{b}(az + b)}{(\bar{b}z + \bar{a})^2}\right|\\ &= \frac{|a|^2 - |b|^2}{|\bar{b}z + \bar{a}|^2}\\ &= \frac{|a|^2 - |b|^2}{(\bar{b}z + \bar{a})(b\bar{z} + a)}.\end{aligned}$$

Thus $(\psi \circ f)(z) \cdot |f'(z)| = \psi(z)$, and so

$$\begin{aligned}\int_{f(\gamma)} \psi(z)|dz| &= \int_0^T (\psi \circ f)(\gamma(t)) \cdot |f'(\gamma(t))| \cdot |\gamma'(t)|dt\\ &= \int_0^T \psi(\gamma(t)) \cdot |\gamma'(t)|dt\\ &= \int_\gamma \psi(z)|dz|.\end{aligned}$$

3. We have to show that $f(\gamma)$ is the SHORTEST curve joining $f(z_1)$ and $f(z_2)$. Let us denote by $L(\delta)$ the LENGTH of a curve δ. We have to show that $L(f(\gamma)) \leqslant L(\delta)$ for any curve δ joining $f(z_1)$ and $f(z_2)$, and that if $L(f(\gamma)) = L(\delta)$ then δ is a reparametrization of $f(\gamma)$. We know that $L(f(\gamma)) = L(\gamma)$, since f preserves LENGTHS; similarly $L(\delta) = L(f^{-1}(\delta))$, and $f^{-1}(\delta)$ is a curve joining z_1 and z_2. But γ is the LINE joining z_1 and z_2. Thus $L(\gamma) \leqslant L(f^{-1}(\delta))$, and equality can hold only if $f^{-1}(\delta)$ is a reparametrization of γ. So $L(f(\gamma)) \leqslant L(\delta)$, and equality can hold only if δ is a reparametrization of $f(\gamma)$.

4. (a) Since $L(O, P) = \dfrac{1}{2}\log\dfrac{1 + r}{1 - r}$, we have

$$\frac{1 + r}{1 - r} = e^{2L(O,P)},$$

from which it immediately follows that

$$r = \tanh L(O, P).$$

A rotation about the origin is an analytic equivalence of the unit disc with itself, and so preserves both Euclidean length and Lobachevskian LENGTH. If Q is a point of the unit disc, we may use a rotation to bring it to a point on the positive real axis, for which $r = OQ$. Thus $OQ = \tanh L(O, Q)$.

(b) We calculate the LENGTH $L(P, Q)$ by using an analytic equivalence of the unit disc with itself to bring P to the origin. If we suppose that P is the point corresponding to the real number p, then $z \longrightarrow \dfrac{z - p}{1 - zp}$ is a suitable mapping. Now if Q is the point corresponding to iq (it lies on the imaginary axis) then Q', the image of Q under this transformation, is $\dfrac{iq - p}{1 - ipq}$.

Now $L(P, Q) = L(O, Q')$, and $\tanh L(O, Q') = OQ'$. But

$$OQ' = \left|\frac{iq - p}{1 - ipq}\right| = \frac{p^2 + q^2}{1 + p^2q^2}.$$

Thus

$$\tanh^2 L(P, Q) = \frac{\tanh^2 L(O, P) + \tanh^2 L(O, Q)}{1 + \tanh^2 L(O, P) \cdot \tanh^2 L(O, Q)}.$$

But $\cosh 2x = \dfrac{1 + \tanh^2 x}{1 - \tanh^2 x}$ for any real number x. So

$$\begin{aligned}
\cosh 2L(P, Q) &= \frac{1 + \tanh^2 L(P, Q)}{1 - \tanh^2 L(P, Q)} \\
&= \frac{\dfrac{\big(1 + \tanh^2 L(O, P)\big)\big(1 + \tanh^2 L(O, Q)\big)}{1 + \tanh^2 L(O, P) \cdot \tanh^2 L(O, Q)}}{\dfrac{\big(1 - \tanh^2 L(O, P)\big)\big(1 - \tanh^2 L(O, Q)\big)}{1 + \tanh^2 L(O, P) \cdot \tanh^2 L(O, Q)}} \\
&= \frac{1 + \tanh^2 L(O, P)}{1 - \tanh^2 L(O, P)} \cdot \frac{1 + \tanh^2 L(O, Q)}{1 - \tanh^2 L(O, Q)} \\
&= \cosh 2L(O, P) \cdot \cosh 2L(O, Q).
\end{aligned}$$

NOTATION

The following items of notation are explained on the pages given. Most other items of notation used in the text may be found in one of the following: *Mathematical Handbook* for M100, The Mathematics Foundation Course; *Handbook* for M231, Analysis; *Units 1, 2, 3, 4, 5, 6, 8, 9, 10 and 11* of M332, Complex Analysis.

$\nabla_\alpha \phi(x, y)$ 15

$\lim\limits_{z \to \infty} f(z) = l$ 45

$\lim\limits_{z \to a} f(z) = \infty$ 45

$\hat{\mathbf{C}} = \mathbf{C} \cup \{\infty\}$ 46

$z = \infty, z \neq \infty$ 46

$\hat{f}$ 46

$\hat{f}(\infty) = a$ 46

$\hat{f}(a) = \infty$ 46

$\lim\limits_{z \to \infty} f(z) = \infty$ 46

INDEX

Abstract geometry, 70
Analytically equivalent regions, 58
Analytic equivalence, 58
Angle between arcs, 8
Axiomatic system, 70

Bilinear transformation, 22
  images of circles and straight lines under, 24, 28
  images of regions under, 30

Closed chain of circles, 39
Closed set in $\hat{\mathbb{C}}$, 46
Compact set, 46
Completing the square, 29
Conformal transformation, 9
Continuous extension of f to $\hat{\mathbb{C}}$, 47
Continuous function $\hat{f}$, 46
Contraction, 22
Cross ratio, 38

Directional derivative, 15

Extended complex plane, 46
Extended complex plane model of Riemann's geometry, 72

Fixed point, 39
Function on $\hat{\mathbb{C}}$, 46

Inverse points with respect to a circle, 24
Inversion in unit circle, 22
Isometry, 23

Lagrange's formula, 16
Laurent series, convergent for $|z| > r$, 54
Length in Lobachevskian geometry, 75
Length in Riemann's geometry, 72
Limit of f near a, 46
Lobachevskian geometry, 71

Magnification, 22
Model of an abstract geometry, 71

Neighbourhood of point at infinity, 46
Non-Euclidean geometry, 71

Open mapping, 10
Open Mapping Theorem, 11
Open set in $\hat{\mathbb{C}}$, 46
Open unit disc model of Lobachevskian geometry, 73

Parallel axiom, 70
Picard's method, 16
Point at infinity, 46
  neighbourhood of, 46

Reflection in real axis, 24
Riemann Mapping Theorem, 60
Riemann's Non-Euclidean geometry, 71
Rotation about origin, 22

Similarity, 24
Simply-connected region in $\hat{\mathbb{C}}$, 55
Singularity of f at ∞, 47
Stereographic projection, 48

Tangent to an arc, 8
Taylor series for inverse function, 16
Translation, 22

Validity condition (for a circle), 25

COMPLEX ANALYSIS

1 Complex Numbers
2 Continuous Functions
3 Differentiation
4 Integration
5 Cauchy's Theorem I
6 Taylor Series
7 NO TEXT
8 Singularities
9 Cauchy's Theorem II
10 The Calculus of Residues
11 Analytic Functions
12 NO TEXT
13 Properties of Analytic Functions
14 Laplace Transforms
15 Number Theory
16 Boundary Value Problems

Course Team

Chairman:	Dr. G. A. Read	Senior Lecturer
Authors:	Dr. P. D. Bacsich	Lecturer
	Dr. M. Crampin	Lecturer
	Mr. N. W. Gowar	Senior Lecturer
	Dr. R. J. Wilson	Lecturer
Editor:	Mr. R. J. Knight	Lecturer
B.B.C.:	Mr. D. Saunders	

With assistance from:

Mr. R. Clamp	B.B.C.
Mr. D. Goldrei	Course Assistant
Mr. H. Hoggan	B.B.C.
Mr. T. Lister	Staff Tutor
Mr. R. J. Margolis	Staff Tutor
Dr. A. R. Meetham	Staff Tutor
Mr. J. E. Phythian	Staff Tutor
Mr. J. Richmond	B.B.C.
Mrs. P. M. Shepheard Rogers	Course Assistant
Dr. C. A. Rowley	Course Assistant
Mr. D. Sargent	Course Assistant

OVE

KM-487-050

0 335